

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 04:03:35 ; Search time 303 Seconds
(without alignments)
4072.921 Million cell updates/sec

Title: US-09-744-679-7
Perfect score: 548
Sequence: 1 cgcgaggtcgtctggacc.....ccatggtgttcatcctg 548

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	548	100.0	548	21	AAZ50772 Human thrombin rec
2	548	100.0	3472	20	AAZ32191 Human thrombin rec
3	548	100.0	3480	13	AAQ28568 Human thrombin rec
4	548	100.0	3480	21	AAZ50771 Human thrombin rec
5	387	70.6	1278	24	ABK70889 Human cDNA encodin
6	328.4	59.9	2910	18	AAT62461 Thr-GPAL fusion ge
7	322.4	58.8	6546	21	AAZ21433 Human low adenosin
8	322.4	58.8	6546	21	AAA35311 Human adenosine re
9	321	58.6	3182	21	AAF21432 Human low adenosin

10	321	58.6	3182	21	AAA35310 Human adenosine re
11	282.4	51.5	1764	15	AAQ73590 Fragment of the hu
12	249	45.4	772	23	AAS68315 DNA encoding novel
13	240.8	43.9	1209	24	ABK70887 Human cDNA encodin
14	184.2	33.6	1312	18	AAT93367 Mouse thrombin rec
15	148.2	27.0	6436	24	ABL32681 Human immune syste
16	147	26.8	1116	24	ABK70888 Human cDNA encodin
17	146.6	26.8	6436	24	ABL32680 Human immune syste
18	146	26.6	572	21	AAA27044 Human cell surface
19	137.2	25.0	556	24	ABQ23658 Oligonucleotide fo
20	137.2	25.0	556	24	ABQ23659 Oligonucleotide fo
21	135.6	24.7	556	24	ABQ23660 Oligonucleotide fo
22	135.6	24.7	556	24	ABQ23661 Oligonucleotide fo
23	126.6	23.1	1300	19	AAV33372 PROLAC FLAG-conne
24	125.4	22.9	1300	19	AAV33371 PROLAC FLAG-conne
25	56	10.2	56	19	AAV33375 Fragment of thromb
26	55	10.0	55	19	AAV33376 Fragment of thromb
27	48.8	8.9	47981	22	AAF30757 Micromonospora meg
28	46	8.4	54	19	AAV33374 Thrombin site-DR b
29	44.6	8.1	192	18	AAV33373 Human nervous syst
30	44.4	8.1	12111	22	ABA21422 Mouse ischaemic co
31	42	7.7	3272	24	ABI99475 DNA encoding novel
32	41.8	7.6	2259	23	AAV79424 WFS1 variant genom
33	40.8	7.4	67212	21	AAQ08954 Thrombin site-DR a
34	40.6	7.4	189	18	AAV97117 HSV-2 strain SB5 C
35	40.6	7.4	2341	19	AAV62155 HSV-2 strain SB5 C
36	40.6	7.4	2694	19	AAV62145 HSV-2 strain SB5 C
37	40.6	7.4	117213	19	AAV62176 Human herpesvirus
38	40.6	7.4	154746	24	AAV25519 CNS disorder-relat
39	40.4	7.4	327	23	AAH88285 Human cardiovascular
40	40.2	7.3	32195	22	AAS36105 Human DNA for a no
41	40.2	7.3	32195	22	AAS31532 Human DNA for a no
42	40.2	7.3	32195	22	AAS31538 Genomic DNA encodi
43	40.2	7.3	32195	23	ABK44045 Human polynucleoti
44	40.2	7.3	32195	24	ABQ66856 Human polynucleoti
45	40.2	7.3	32195	24	ABQ66862

ALIGNMENTS

RESULT 1
AAZ50772
ID AAZ50772 standard; cDNA; 548 BP.
XX AAZ50772;
AC AAZ50772;
XX AAZ50772;
DT 31-MAY-2000 (first entry)
XX Human thrombin receptor antisense cDNA.
DE Human; thrombin receptor; THR; PAR-1; protease activated receptor;
KW antisense molecule; PAR antibody; cytostatic; therapeutic;
KW metastatic tumour cell; placental implantation; invasive cell; ss.
XX Homo sapiens.
OS Homo sapiens.
XX WO2000008150-A1.
PN WO2000008150-A1.
XX 17-FEB-2000.
PD 17-FEB-2000.
XX 05-FEB-1999; 99WO-IL00079.
PF 05-FEB-1999; 99WO-IL00079.
XX 07-AUG-1998; 98IL-0125698.
PR (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
PA Bar-Shavit R;
PI Bar-Shavit R;
XX WPI; 2000-205706/18.
DR Treating metastatic tumor cells useful for treating disorders involving
XX Placenta implantation in a female comprises administration of an
PT

polymorphic site; phenotype correlation; forensic; paternity testing; medicine; genetic analysis; vascular disease; ds.

antisenase molecule complementary to an RNA sequence of a protease activated receptor protein -

Claim 6; Fig 2; 46pp; English.

The patent discloses a method to treat metastatic tumour cells using an antisenase molecule comprising a polynucleotide complementary to an RNA sequence of a protease activated receptor (PAR) protein, or an antibody capable of binding to a PAR protein. The antisenase molecules and antibodies of PAR protein are also used to treat disorders associated with implantation of placenta. The present sequence is an antisenase cDNA of human thrombin receptor (THR) (also known as PAR-1), which is a G-coupled protein belonging to the PAR family. This sequence is useful in the treatment of invasive cells.

Sequence 548 BP; 122 A; 166 C; 148 G; 112 T; 0 other;

Query Match 100.0%; Score 548; DB 21; Length 548;
Best Local Similarity 100.0%; Pred. No. 6.8e-139;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CGCGAGGGTGGACCTGATCTTACCGTGGGCACCCCTGCGCTCTGCTGCGCGG 60
1 CGCGAGGGTGGACCTGATCTTACCGTGGGCACCCCTGCGCTCTGCTGCGCGG 60

61 AAGACCGGCTCCCGACCCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
61 AAGACCGGCTCCCGACCCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

121 GGGGAGGCTCCCGAG 180
121 GGGGAGGCTCCCGAG 180

181 TGCTGGTGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
181 TGCTGGTGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240

241 CGAGGCCAGAGTCAAAAGCAAAATGACCTTACCTTACCTTACCTTACCTTACCT 300
241 CGAGGCCAGAGTCAAAAGCAAAATGACCTTACCTTACCTTACCTTACCTTACCT 300

301 ACCCAATGATAATATGACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 360
301 ACCCAATGATAATATGACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 360

361 CTGAATACAGATTAGTCTCCATCAATCAATCAATCAATCAATCAATCAATCAATCA 420
361 CTGAATACAGATTAGTCTCCATCAATCAATCAATCAATCAATCAATCAATCAATCA 420

421 TCATCTCAGAAATGCTCCGAGATTTGACCAAGTCTGCTGCTGCTGCTGCTGCTGCT 480
421 TCATCTCAGAAATGCTCCGAGATTTGACCAAGTCTGCTGCTGCTGCTGCTGCTGCT 480

481 CTGTGTACACCGAGGTGTTGTAGTACGCTCCCACTAAACATCATGCGCATCGTTGT 540
481 CTGTGTACACCGAGGTGTTGTAGTACGCTCCCACTAAACATCATGCGCATCGTTGT 540

541 TCATCTCTG 548
541 TCATCTCTG 548

RESULT 2
AAZ32191 standard; cDNA; 3472 BP.

AAZ32191;
13-JAN-2000 (first entry)
Human thrombin receptor nucleotide sequence.
Human; coding sequence polymorphism; vascular pathology gene.

KW
KW
XX
OS
XX
PN
XX
XX
PD
XX
PF
XX
PR
XX
XX
PA
XX
PI
XX
DR
XX
XX
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Homo sapiens.
WO9950454-A2.
07-OCT-1999.
26-MAR-1999; 99WO-US06473.
01-APR-1998; 98US-0054272.
(WHED) WHITEHEAD INST BIOMEDICAL RES.
Lander ES, Daley GO, Cargill M, Ireland JS, Rozen SG;
WPI; 1999-620066/53.
P-PSDB; AAY49570.
Determination of polymorphisms in genes, especially those identifying predisposition to vascular disease -
Claim 1; Fig 35; 134pp; English.
AAZ32159 to AAZ32194 represent reference alleles for specifically claimed nucleic acid sequences from the present invention which comprise polymorphic sites as given in a table in the specification, selected from 92 single nucleotide polymorphisms in which the nucleotide at the polymorphic site is different from a nucleotide at the same site in a reference allele. The nucleic acids, and primers and probes, are used to identify polymorphisms, which may predispose an individual to disease, especially a vascular disease. They can also be used in phenotypic correlations, forensics, paternity testing, medicine or genetic analysis. AAY49550 to AAY49573 represent the proteins which correspond to some of the reference alleles.

Sequence 3472 BP; 933 A; 817 C; 785 G; 937 T; 0 other;

Query Match 100.0%; Score 548; DB 20; Length 3472;
Best Local Similarity 100.0%; Pred. No. 1.4e-138;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CGCGAGGGTGGACCTGATCTTACCGTGGGCACCCCTGCGCTCTGCTGCGCGG 60
64 CGCGAGGGTGGACCTGATCTTACCGTGGGCACCCCTGCGCTCTGCTGCGCGG 123

61 AAGACCGGCTCCCGACCCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
124 AAGACCGGCTCCCGACCCGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 183

121 GGGGAGGCTCCCGAG 180
184 GGGGAGGCTCCCGAG 243

181 TGCTGGTGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
244 TGCTGGTGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 303

241 CGAGGCCAGAGTCAAAAGCAAAATGACCTTACCTTACCTTACCTTACCTTACCT 300
304 CGAGGCCAGAGTCAAAAGCAAAATGACCTTACCTTACCTTACCTTACCTTACCT 363

301 ACCCAATGATAATATGACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 360
364 ACCCAATGATAATATGACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCT 423

361 CTGAATACAGATTAGTCTCCATCAATCAATCAATCAATCAATCAATCAATCAATCA 420
424 CTGAATACAGATTAGTCTCCATCAATCAATCAATCAATCAATCAATCAATCAATCA 483

421 TCATCTCAGAAATGCTCCGAGATTTGACCAAGTCTGCTGCTGCTGCTGCTGCTGCT 480

```

Db 484 TCATCTCAGAAAGATCCCTCCGGATATTGACACAGCTCCTGGCTGACACTCTTGTCCCAT 543
Qy 481 CTGTGTACACCGGAGTGTTCGTAGTCAAGCTCCACCTAAACATATGCGCATCGTGTGTGT 540
Db 544 CTGTGTACACCGGAGTGTTCGTAGTCAAGCTCCACCTAAACATATGCGCATCGTGTGTGT 603
Qy 541 TCATCCTTG 548
Db 604 TCATCCTG 611

RESULT 3
AAQ28568
ID AAQ28568 standard; DNA; 3480 BP.
XX
AC AAQ28568;
DT 15-FEB-1993 (first entry)
DE Human thrombin receptor gene.
KW Diagnosis; cardiovascular disease; wound healing; restenosis;
KW thrombosis; unstable angina treatment; myocardial infarction;
KW thrombotic; thromboembolytic stroke; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 225..1502
FT sig_peptide /*tag= a
FT mat_peptide 225..303
FT 304..1502
FT /*tag= c
XX
PN WO9214750-A.
PD 03-SEP-1992.
PF 19-FEB-1992; 92WO-US01312.
PR 19-FEB-1991; 91US-0657769.
PR 07-NOV-1991; 91US-0789184.
PA (CORT-) COR THERAPEUTICS INC.
PA (REGC) UNIV CALIFORNIA.
PI Coughlin SR, Scarborough RM;
XX
DR WPI: 1992-316119/38.
DR P-PSDB; AAR27240.
XX
FT DNA encoding cell surface receptor for thrombin - useful for
FT determining thrombin in diagnosing e.g. cardiovascular diseases,
FT also to treat wound healing, restenosis etc.
XX
PS Disclosure; Fig 1; 8lpp; English.
XX
CC The sequence is that of the gene coding for human thrombin receptor.
CC It can be used in the prepn. of diagnostics to determine thrombin
CC levels in samples, and screening tools for candidate substances which
CC affect thrombin activity in vivo. Thrombosis may be diagnosed in a
CC mammal by measuring the presence, absence or amt. of the cleaved
CC activation peptide of the TR.
XX
SQ Sequence 3480 BP; 947 A; 816 C; 786 G; 931 T; 0 other;

Query Match 100.0%; Score 548; DB 13; Length 3480;
Best Local Similarity 100.0%; Pred. No. 1.4e-138;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCCGAGGGTCGCTTGACCTGATCTTACCGTGGCAGCCCTCGCTCGCTGCGCG 60
|||||

```

```

Db 64 GCGCCGAGGTCGCTTGGACCTGATCTTACCGTGGCACCCTGCTGCTGCTGCCGG 123
Qy 61 AAGACCCGCTCCCGACCCGAGAAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 120
Db 124 AAGACCCGCTCCCGACCCGAGAAAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCCGAGGC 183
Qy 121 GGGCAGCCTCCCGAGAGCGCCGCGAGAGCCCGGAGCAATGGGGCGGCGGCTGC 180
Db 184 GGGCAGCCTCCCGAGAGCGCCGCGAGAGCCCGGAGCAATGGGGCGGCGGCTGC 243
Qy 181 TGCTGGTGGCGCTGCTTCACTCTGTCGCGCCGCTGTTGTCGCCGCGACCCGGGCCC 240
Db 244 TGTGTGGTGGCGCTGCTTCACTCTGTCGCGCCGCTGTTGTCGCCGCGACCCGGGCCC 303
Qy 241 GCAGGCCAGATCAAAAGCAAAATGCCACTTTAGATCCCGGTCATTTCTTCTCAGGA 300
Db 304 GCAGGCCAGATCAAAAGCAAAATGCCACTTTAGATCCCGGTCATTTCTTCTCAGGA 363
Qy 301 ACCCAATGATAATATGACCAATTTGGGAGGATGAGGAGAAAAATCAAAAGTGGTTAA 360
Db 364 ACCCAATGATAATATGACCAATTTGGGAGGATGAGGAGAAAAATCAAAAGTGGTTAA 423
Qy 361 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCTCTTCAAAAACAATTCCTGCAT 420
Db 424 CTGAATACAGATTAGTCTCCATCAATAAAAGCAGTCTCTTCAAAAACAATTCCTGCAT 483
Qy 421 TCATCTCAGAAAGTGCCTCCGGATATTGACCACTCCTGGCTGACACTCTTTGTCCTCAT 480
Db 484 TCATCTCAGAAAGTGCCTCCGGATATTGACCACTCCTGGCTGACACTCTTTGTCCTCAT 543
Qy 481 CTGTGTACACCGGAGTGTTCGTAGTCAAGCTCCAGCTCCCAACATCATGGCCATCGTTGTGT 540
Db 544 CTGTGTACACCGGAGTGTTCGTAGTCAAGCTCCAGCTCCCAACATCATGGCCATCGTTGTGT 603
Qy 541 TCATCCTG 548
Db 604 TCATCCTG 611

RESULT 4
AAZ50771
ID AAZ50771 standard; DNA; 3480 BP.
XX
AC AAZ50771;
DT 31-MAY-2000 (first entry)
DE Human thrombin receptor DNA.
XX
KW Human; thrombin receptor; ThR; PAR-1; protease activated receptor;
KW antiseize molecule; PAR antibody; cytostatic; therapeutic;
KW metastatic tumour cell; placental implantation; invasive cell; ds.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 225..1502
FT /*tag= a
FT /product= "Thrombin receptor"
FT /transl_except= (pos:1500..1502, aa:Glx)
FT /note= "no stop codon given"
FT misc_binding 320..570
FT /*tag= b
FT /bound_moiety= "ThR RNA probe"
XX
PN WO200008150-A1.
XX
PD 17-FEB-2000.
XX
PF 05-FEB-1999; 99WO-IL00079.
XX
PR 07-AUG-1998; 98IL-0125698.
XX

```

PA	(HADA-) HADASIT MEDICAL RES SERVICES & DEV.	ABK70889
XX	Bar-Shavit R;	ID ABK70889 standard; cDNA; 1278 BP.
PI		XX
XX		AC ABK70889;
DR	WPI; 2000-205706/18.	XX
DR	P-PSDB; AAY45035.	DT 15-JUL-2002 (first entry)
XX	Treating metastatic tumor cells useful for treating disorders involving	XX
PT	placenta implantation in a female comprises administration of an	DE Human cDNA encoding PAR1 type thrombin receptor.
PT	antisense molecule complementary to an RNA sequence of a protease	XX Human; ss; gene; PAR1; thrombin receptor; antinflammatory; cytostatic;
PT	activated receptor protein	KW inflammatory disease; cell proliferative disease.
XX		XX Homo sapiens.
PS	Example 2; Fig 1a; 46pp; English.	XX
XX	The patent discloses a method to treat metastatic tumour cells using	PN JP2002010784-A.
CC	an antisense molecule comprising a polynucleotide complementary to an	XX
CC	RNA sequence of a protease activated receptor (PAR) protein, or an	PD 15-JAN-2002.
CC	antibody capable of binding to a PAR protein. The antisense molecules and	XX
CC	antibodies of PAR protein are also used to treat disorders associated	PF 29-JUN-2000; 2000JP-0196514.
CC	with implantation of placenta. The present sequence is a DNA encoding	XX
CC	human thrombin receptor (ThR) (also known as PAR-1), which is a	PR 29-JUN-2000; 2000JP-0196514.
CC	G-coupled protein belonging to the PAR family. This sequence is used to	XX (TEIJ) TEIJIN LTD.
CC	produce antisense molecules or antibodies against PAR protein for	PA
CC	therapeutic use.	XX
XX		DR WPI; 2002-321520/36.
SQ	Sequence 3480 BP; 941 A; 817 C; 785 G; 937 T; 0 other;	P-PSDB; ABG35300.
	Query Match 100.0%; Score 548; DB 21; Length 3480;	XX
	Best Local Similarity 100.0%; Pred. No. 1.4e-138;	PT An inhibitor of cell growth mediated by thrombin used to treat
	Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	PT inflammatory and cell proliferative diseases -
		XX
QY	1 GCCGAGGGTGGACCTGATCTTACCCGTGGCACCCTCGCTCGCTGCGCGG 60	PS Claim 9; Page 26-27; 44pp; Japanese.
Db	64 GCCGAGGGTGGACCTGATCTTACCCGTGGCACCCTCGCTCGCTGCGCGG 123	XX The invention relates to a polypeptide or a compound which can inhibit
QY	61 AGACCGGCTCCCGACCCGACAGGTGAGGAGAGAGGTGAAGCGGAGCGCCGAGGC 120	CC cell growth caused by thrombin. The polypeptide/compound combines to a
Db	124 AGACCGGCTCCCGACCCGACAGGTGAGGAGAGAGGTGAAGCGGAGCGCCGAGGC 183	CC specific region of the structure of PAR1 type human thrombin receptor
QY	121 GGGGAGGCTCCCGAGGAGCGCGGAGAGCGCGGAGAGGTGAAGCGGAGCGCCGAGGC 180	CC participating to cell growth. Preferably, the compound contains the
Db	184 GGGGAGGCTCCCGAGGAGCGCGGAGAGCGCGGAGAGGTGAAGCGGAGCGCCGAGGC 243	CC 52nd to the 56th amino acid sequences at the amino end side of PAR1 type
QY	181 TGCTGGTGGCCGCTTTCAGTCTGTGCGGCCGCTGTGTCTGCGCCGACCCGCGGCC 240	CC human thrombin receptor ((X4)-Tyr-Glu-Pro-Phe-Trp-(X5) X4, X5 = optional
Db	244 TGCTGGTGGCCGCTTTCAGTCTGTGCGGCCGCTGTGTCTGCGCCGACCCGCGGCC 303	CC amino acid or peptide sequence). Also included are a modified PAR1 type
QY	241 GCAGCCAGAAATCAAAAGCAACAAATGCCATTCAGTCCCGGTCATTTCTTCTCAGGA 300	CC thrombin receptor gene or its fragment used for obtaining the above
Db	304 GCAGCCAGAAATCAAAAGCAACAAATGCCATTCAGTCCCGGTCATTTCTTCTCAGGA 363	CC polypeptide, a human PAR1 type thrombin receptor protein and its
QY	301 ACCCCNATGATAAATATGAAACCATTTTGGAGGATGAGAGAAAATGAAGTGGTTAA 360	CC encoding DNA comprising a fully. The polypeptide or the compound is used
Db	364 ACCCCNATGATAAATATGAAACCATTTTGGAGGATGAGAGAAAATGAAGTGGTTAA 423	CC to treat inflammatory diseases and cell proliferative diseases. The
QY	361 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACCTTCCTGAT 420	CC present sequence is a cDNA encoding a PAR1 type thrombin receptor (or a
Db	424 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACCTTCCTGAT 483	CC modified version).
QY	421 TCATCTCAGAAGATGCTCCGATATTTGACCACTCTCGGCTGACACTTTTGTCCCAT 480	XX Sequence 1278 BP; 276 A; 353 C; 284 G; 365 T; 0 other;
Db	484 TCATCTCAGAAGATGCTCCGATATTTGACCACTCTCGGCTGACACTTTTGTCCCAT 543	Query Match 70.6%; Score 387; DB 24; Length 1278;
QY	481 CTGTGTACACCGAGTGTGTGTAGTCAGGCTCCCACTAAACATCATGGCCATCTGTGT 540	Best Local Similarity 100.0%; Pred. No. 4.8e-95;
Db	544 CTGTGTACACCGAGTGTGTGTAGTCAGGCTCCCACTAAACATCATGGCCATCTGTGT 603	Matches 387; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	541 TCATCTCTG 548	QY 162 ATGGGGCGCGGGCGCTGCTGCTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 221
Db	604 TCATCTCTG 611	Db 1 ATGGGGCGCGGGCGCTGCTGCTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
RESULT	5	QY 222 TCTGCCCGACCCCGGGCGCGCGAGATCAAAAGCAACAAATGCCATTCAGTATGCC 281
		Db 61 TCTGCCCGACCCCGGGCGCGCGAGATCAAAAGCAACAAATGCCATTCAGTATGCC 120
		QY 282 CGGTCAATTTCTTCAGAAACCCCAATGATAATATGAACCATTTTGGGAGGATGAGGAG 341
		Db 121 CGGTCAATTTCTTCAGAAACCCCAATGATAATATGAACCATTTTGGGAGGATGAGGAG 180
		QY 342 AAAAATGAAAGTGGTAACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTT 401
		Db 181 AAAAATGAAAGTGGTAACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTT 240
		QY 402 CAAAACAACTTCTGCAATTCATCTCAGAGATGCCCTCCCGATATTTGACAGCTCTGG 461
		Db 241 CAAAACAACTTCTGCAATTCATCTCAGAGATGCCCTCCCGATATTTGACAGCTCTGG 300
		QY 462 CTGACACTCTTTGTGCCCATCTGTGTACACCGAGTGTGTGTAGTCAGCTCCCACTAAAC 521

Db 301 CTGACACTCTTTGTCCCATCTGTGTACCGGAGTGTTTGTAGTCAGCCTCCCACTAAAC 360

QY 522 ATCATGGCCATCGTTGTGTTTCACCTG 548

Db 361 ATCATGGCCATCGTTGTGTTTCACCTG 387

RESULT 6	
AY62461	
ID	AAT62461 standard; DNA; 2910 BP.
XX	
AC	AAT62461;
XX	
DT	16-AUG-1997 (first entry)
XX	
DE	Thr-GPAL fusion gene.
XX	
KW	G-protein coupled receptor; agonist; antagonist; assay;
KW	G-alpha protein; Gpalp; GPAL gene; thrombin receptor; ds.
XX	
OS	Chimaeric Homo sapiens;
OS	Chimaeric Saccharomyces cerevisiae.
XX	
Key	Location/Qualifiers
FT	1..2679
CDS	
FT	/*tag= a
FT	/product= Thr-GPAL fusion
FT	1..39
FT	/*tag= b
FT	/note= "STET nucleotides 535-573"
FT	40..1251
FT	/*tag= c
FT	/product= thrombin receptor
FT	/note= "corresponds to nucleotides 288-1499 of
FT	the thrombin receptor gene"
FT	1252..1260
FT	/*tag= d
FT	/note= "STET nucleotides 1827..1835"
FT	1261..2910
FT	/*tag= e
FT	/product= Gpalp
FT	/note= "corresponds to nucleotides 202-1851 of
FT	GPAL"
FT	complement (42..62)
FT	/*tag= f
FT	/note= "oligonucleotide g for thrombin receptor
FT	gene amplification"
FT	1229..1251
FT	/*tag= g
FT	/note= "oligonucleotide h for thrombin receptor
FT	gene amplification"
XX	
XX	W09711159-A1.
XX	
XX	27-MAR-1997.
XX	
XX	20-SEP-1996; 96WO-US15203.
PPF	
XX	
XX	20-SEP-1995; 95US-0004023.
PPR	
XX	
XX	(HEAR-) HEARTLAND BIOTECHNOLOGIES LLC.
XX	
XX	Boulton TG, Das PO, Mandell RB, McMullen TW;
PI	
XX	WPI; 1997-202868/18.
OR	P-PSDB; AAW16314.

[illegible]

XX PR 06-APR-1999; 99US-0127958.
XX PA (UYEC-) UNIV EAST CAROLINA.
XX PI (NYCE/) NYCE J W.
XX NYce JW;
XX WPI; 2000-679539/66.
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
XX trigger adenosine receptors during metabolism, useful e.g. for treating
XX cancers and respiratory obstructions -
XX
XX Disclosure; Page 1416-1418; 1592pp; English.
XX
XX The present invention describes low adenosine (A) content antisense
XX oligonucleotides and compositions (I) comprising them. In the antisense
XX oligonucleotides the A is replaced by a 'Universal' or alternative base.
XX oligonucleotides, bronchodilator, antiinflammatory, analgesic,
XX (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
XX immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
XX The antisense oligonucleotides and (I) can be used to down-regulate the
XX expression and/or activity of target polypeptides associated with
XX lung/respiratory disorders and malignancies, such as stimulating factors,
XX activating peptide factors and transmitters, transcription factors and
XX immunoglobulins and antibodies, antibody receptors, cytokines and
XX chemokines, endogenously produced specific and non-specific enzymes,
XX binding proteins, adhesion molecules and their receptors, cytokine and
XX chemokine receptors, adenosine receptors, bradykinin receptors, central
XX nervous system (CNS) and peripheral nervous and non-nervous system
XX receptors, CNS and peripheral nervous and non-nervous system peptide
XX transmitters, defensins, growth factors, vasoactive peptides and
XX receptors, binding proteins and malignancy associated proteins. The
XX antisense oligonucleotides may be used in this way to treat disorders
XX including respiratory obstruction (especially pulmonary obstruction
XX and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
XX and/or surfactant hypoproduction which are associated with a disease or
XX condition selected from pulmonary vasoconstriction, inflammation,
XX allergies, asthma, impeded respiration, respiratory distress syndrome
XX (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
XX hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
XX pulmonary transplantation rejection, pulmonary infections, bronchitis,
XX and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
XX fragments and antisense oligonucleotides used in the exemplification of
XX the present invention.
XX
XX Sequence 6546 BP; 1603 A; 1691 C; 1500 G; 1751 T; 1 other;

Db 3662 ATGGCCATCGTTGTTCATCTG 3685
|||||
RESULT 8
AAA35311
ID AAA35311 standard; DNA; 6546 BP.
XX AC AAA35311;
XX 28-JUL-2000 (first entry)
XX Human adenosine receptor related polynucleotide 2nd SEQ ID NO:185.
XX
XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
XX phosphorothioate; impaired respiration; inflammation; allergy;
XX allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
XX antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
XX lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
XX respiratory distress syndrome; pain; cystic fibrosis; emphysema;
XX pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
XX cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
XX Homo sapiens.
OS
XX WO200009525-A2.
PN
XX 24-FEB-2000.
PD
XX 03-AUG-1999; 99WO-US17712.
PF
XX 03-AUG-1998; 98US-0095212.
PR
XX (UYEC-) UNIV EAST CAROLINA.
PA
XX Nyce JW;
XX WPI; 2000-205971/18.
DR
XX New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasoconstriction, inflammation, allergies, asthma, hypertension,
XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
XX cancers -
XX
XX Disclosure; Page 1331-1333; 1343pp; English.
XX
XX The present invention describes a new composition comprising an
XX antisense oligonucleotide (ON) with low adenosine (up to 15%), which
XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
XX inflammation. The ON can have antiinflammatory, antiallergic,
XX antiasthmatic, cytostatic and analgesic activities. The compositions are
XX useful for the treatment of diseases associated with inflammation,
XX impaired airways, including lung disease and diseases whose secondary
XX effects afflict the lungs of a subject. They can be used for treating
XX e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
XX asthma, impeded respiration, respiratory distress syndrome, pain, cystic
XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
XX carcinomas, and cancers which may metastasise to the lungs, including
XX breast and prostate cancer. The reduction of the adenosine content of
XX the ONs reduces side effects. The A-containing ONs break down with the
XX release of deoxyadenosine which activates adenosine receptors causing
XX bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
XX nucleotide sequences given in the sequence listing from the present
XX invention, which correspond to SEQ ID NO:1 to 185, and then the last
XX 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
XX differ from the previously named sequences. SEQ ID NO:11 to 1680
XX (AAA32323 to AAA33992) are specifically claimed ONs from the present
XX invention. N.B. Sequences given in the disclosure of the present
XX invention do not match up with their corresponding SEQ ID NO: sequences
XX given in the sequence listing.
XX
XX Sequence 6546 BP; 1603 A; 1692 C; 1500 G; 1750 T; 1 other;

Query Match 58.8%; Score 322.4; DB 21; Length 6546;
 Best Local Similarity 99.7%; Pred. No. 3.1e-77;
 Matches 323; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 225 GCCCGCACCCGGCCCGCAGGCGCAGATCAAAAGCAACAAATGCCACCTTAGATCCCGG 284
 DB 3362 GCACGACCCGGCCCGCAGGCGCAGATCAAAAGCAACAAATGCCACCTTAGATCCCGG 3421

QY 285 TCATTTCTTCTCAGGAACCCCAATGATAATATGAACCATTTTGGGAGGATGAGGAGAA 344
 DB 3422 TCATTTCTTCTCAGGAACCCCAATGATAATATGAACCATTTTGGGAGGATGAGGAGAA 3481

QY 345 ATGAAAGTGGGTAACTGAATACAGATTAGTCTCATCAATAAAGAGCAGTCTCTTCAA 404
 DB 3482 AATGAAAGTGGGTAACTGAATACAGATTAGTCTCATCAATAAAGAGCAGTCTCTTCAA 3541

QY 405 AAACAACTTCCTGCATTCATCTCAGAGATGCTCCGGATATTTGACAGCTCCTGGCTG 464
 DB 3542 AAACAACTTCCTGCATTCATCTCAGAGATGCTCCGGATATTTGACAGCTCCTGGCTG 3601

QY 465 ACACCTTTTGTCCCATCTGTGTACACCGGAGTGTGTGTAGTCAGCTCCCACTAAACATC 524
 DB 3602 ACACCTTTTGTCCCATCTGTGTACACCGGAGTGTGTGTAGTCAGCTCCCACTAAACATC 3661

QY 525 ATGGCCATCGTTGTTTCATCCTG 548
 DB 3662 ATGGCCATCGTTGTTTCATCCTG 3685

RESULT 9
 AAF21432
 ID AAF21432 standard; DNA; 3182 BP.
 XX AAF21432;
 AC XX
 XX XX
 DT 14-MAR-2001 (first entry)
 XX XX
 DE Human low adenosine antisense oligonucleotide related sequence #2999.
 XX XX
 KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
 KW human; airway disorder; bronchoconstriction; lung inflammation;
 KW surfactant depletion; respiratory; bronchodilator; anti-inflammatory;
 KW immunosuppressive; antihistaminic; analgesic; hypotensive; cytostatic;
 KW respiratory obstruction; pulmonary vasoconstriction; impeded respiration;
 KW surfactant hypoproduction; pulmonary obstruction; asthma; RDS;
 KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
 KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
 KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
 KW cancer; ss.
 XX XX
 OS Homo sapiens.
 XX XX
 -FN W020062736-A2.
 XX XX
 PD 26-OCT-2000.
 XX XX
 PF 24-MAR-2000; 2000WO-US08020.
 XX XX
 PR 06-APR-1999; 9905-0127958.
 XX XX
 PA (UYEC-) UNIV EAST CAROLINA.
 PA (NYCE/) NYCE J W.
 XX XX
 PI Nyce JW;
 XX XX
 DR WPI; 2000-679539/66.
 XX XX
 PT Low adenosine (A) content antisense oligonucleotides which do not
 PT trigger adenosine receptors during metabolism, useful e.g. for treating
 PT cancers and respiratory obstructions -
 XX XX
 PS Disclosure; Page 1415-1416; 1592pp; English.

XX
 CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antihistaminic, hypotensive and cytostatic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impeded respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.
 XX
 SQ Sequence 3182 BP; 895 A; 701 C; 684 G; 902 T; 0 other;
 Query Match 58.6%; Score 321; DB 21; Length 3182;
 Best Local Similarity 100.0%; Pred. No. 5.6e-77;
 Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 228 CCGACCCCGGGCCCGCAGGCGCAGATCAAAAGCAACAAATGCCACCTTAGATCCCGGTC 287
 DB 1 CCGACCCCGGGCCCGCAGGCGCAGATCAAAAGCAACAAATGCCACCTTAGATCCCGGTC 60

QY 288 TTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAT 347
 DB 61 TTCTTCTCAGGAACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAT 120

QY 348 GAAAGTGGGTAACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAA 407
 DB 121 GAAAGTGGGTAACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAA 180

QY 408 CAACCTTCCTGATTCATCTCAGAGATGCTCCCGATATTTGACAGCTCCTGGGTGACA 467
 DB 181 CAACCTTCCTGATTCATCTCAGAGATGCTCCCGATATTTGACAGCTCCTGGGTGACA 240

QY 468 CTCCTTTGCCCATCTGTGTACACCGGAGTGTGTGTAGTCAGCTCCCACTAAACATCATG 527
 DB 241 CTCCTTTGCCCATCTGTGTACACCGGAGTGTGTGTAGTCAGCTCCCACTAAACATCATG 300

QY 528 GCCATCGTTGTTCATCCTG 548
 DB 301 GCCATCGTTGTTCATCCTG 321

RESULT 10
 AAA35310
 ID AAA35310 standard; DNA; 3182 BP.
 XX AAA35310;
 AC AAA35310;
 XX XX
 DT 28-JUL-2000 (first entry)
 XX XX
 DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:184.
 XX XX
 KW Human; adenosine receptor; low adenosine antisense oligonucleotide;

KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
XX WO200009525-A2.
XX
XX 24-FEB-2000.
XX
XX 03-AUG-1999; 99WO-US17712.
XX
XX 03-AUG-1998; 98US-0095212.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX
XX Nyce JW;
XX
XX WPI; 2000-205971/18.
XX
XX New antisense oligonucleotides useful for treating e.g. pulmonary
XX vasoconstriction, inflammation, allergies, asthma, hypertension,
XX bronchitis, emphysema, respiratory distress syndrome, ischemia or
XX cancers -
XX
XX Disclosure; Page 1330-1331; 1343pp; English.
XX
XX The present invention describes a new composition comprising an
XX antisense oligonucleotide (ON) with low adenosine (up to 15%), which
XX targets nucleic acids involved in bronchoconstriction, allergies, and/or
XX inflammation. The ON can have antiinflammatory, antiallergic,
XX antiasthmatic, cytostatic and analgesic activities. The compositions are
XX useful for the treatment of diseases associated with inflammation,
XX impaired airways, including lung disease and diseases whose secondary
XX effects afflict the lungs of a subject. They can be used for treating
XX e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
XX asthma, impaired respiration, respiratory distress syndrome, pain, cystic
XX fibrosis, pulmonary hypertension, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
XX carcinomas, and cancers which may metastasize to the lungs, including
XX breast and prostate cancer. The reduction of the adenosine content of
XX the ONs reduces side effects. The A-containing ONs break down with the
XX release of deoxyadenosine which activates adenosine receptors causing
XX bronchoconstriction and inflammation. AAA32313 to AAA3312 represent the
XX nucleotide sequences given in the sequence listing from the present
XX invention, which correspond to SEQ ID NO:1 to 2815, and then the last
XX 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
XX differ from the previously named sequences. SEQ ID NO:11 to 1680
XX (AAA32323 to AAA33992) are specifically claimed ONs from the present
XX invention. N.B. Sequences given in the disclosure of the present
XX invention do not match up with their corresponding SEQ ID NO: sequences
XX given in the sequence listing.
XX
XX Sequence 3182 BP; 895 A; 701 C; 684 G; 902 T; 0 other;
XX
XX Query Match 58.68; Score 321; DB 21; Length 3182;
XX Best Local Similarity 100.0%; Pred. No. 5.6e-77;
XX Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 228 CGCACCCGGCCCGCAGAGCAACAAAGCAACAAATGCCACCTTAGATCCCGGTCA 287
XX 1 CGCACCCGGCCCGCAGAGCAACAAAGCAACAAATGCCACCTTAGATCCCGGTCA 60
XX
XX 288 TTCTTCTCAGAACCCCAATGATAATATGAACCATTTTGGGAGGATGAGGAGAAAT 347
XX 61 TTCTTCTCAGAACCCCAATGATAATATGAACCATTTTGGGAGGATGAGGAGAAAT 120
XX
XX 348 GAAAGTGGGTAACTGAATACAGATTACTCTCCATCAATAAAGCAGTCCTTCAAAA 407
XX
XX

Db 121 GAAAGTGGGTAACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAA 180
QY 408 CAATCTCTGCATTCATCTCAGAGATGCTCCGGATATTTTGACAGCTCCTGGCTGACA 467
Db 181 CAATCTCTGCATTCATCTCAGAGATGCTCCGGATATTTTGACAGCTCCTGGCTGACA 240
QY 468 CTCTTTTCCCATCTGTGTACACGGAGTGTGTGTAGTACAGCTCCACATAAACATCATG 527
Db 241 CTCTTTTCCCATCTGTGTACACGGAGTGTGTGTAGTACAGCTCCACATAAACATCATG 300
QY 528 GCCATCGTTGTGTTCATCCTG 548
Db 301 GCCATCGTTGTGTTCATCCTG 321
XX
XX AAQ73590;
XX
XX 25-JUN-1995 (first entry)
XX
XX Fragment of the human thrombin receptor gene.
XX
XX TR; expression; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 376..1653
XX FT /*tag= a
XX
XX WO9421789-A.
XX
XX 29-SEP-1994.
XX
XX 28-FEB-1994; 94WO-US02388.
XX
XX 25-MAR-1993; 93US-0038662.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Betlach MC, Turner GJ;
XX
XX WPI; 1994-317010/39.
XX
XX P-PSDB; AAR60698.
XX
XX Expression of heterologous proteins in halo-bacteria - using
XX regulatory and stop sequences from halo-bacteria, pref. the
XX bacterio-rhodopsin gene.
XX
XX Disclosure; Fig 14; 118pp; English.
XX
XX The sequence is that of the human thrombin receptor gene fragment.
XX The gene is used to exemplify a new expression vector for producing
XX heterologous polypeptides in a halobacterial host.
XX
XX See also AAQ73586-92.
XX
XX Sequence 1764 BP; 382 A; 492 C; 421 G; 469 T; 0 other;
XX
XX Query Match 51.5%; Score 282.4; DB 15; Length 1764;
XX Best Local Similarity 99.6%; Pred. No. 1.3e-66;
XX Matches 283; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 265 ATGCCACCTTAGATCCCGGTCAATTTCTTCTCAGGAACCCCAATGATAATATGAACCAT 324
XX 461 ACGCCACCTTAGATCCCGGTCAATTTCTTCTCAGGAACCCCAATGATAATATGAACCAT 520
XX
XX 325 TTTTGGGAGGATGAGGAGAAATGAAAGTGGGTAACTGAATACAGATTAGTCTCCATCA 384
XX
XX 521 TTTTGGGAGGATGAGGAGAAATGAAAGTGGGTAACTGAATACAGATTAGTCTCCATCA 580

The present sequence was used in the preparation of a transgenic non-human vertebrate with somatic and germ cells containing an altered form of a thrombin receptor (TR) gene, where the altered gene has replaced a wild type TR gene in the animal or an ancestor at an embryonic stage using embryonic stem cells. The transgenic animal provides an animal model to understand receptor function in, e.g. platelets, endothelial cells and neurons, and evaluate drug therapies modulating TR function/expression in human cells. Alterations rendering the wild type gene non-functional produce knockout animals, useful as non-human whole animal models for diseases involving TR equivalents in humans. Alternatively, substituting the naturally occurring gene for a gene from a 2nd species or with a mutation, results in animals producing the 2nd species or mutated gene products. These transgenic animals are useful for drug antagonist and agonist studies, creation of animal models of human diseases and investigation of disorders associated with human TR-mediated responses. Antagonists may be used therapeutically for cardiovascular problems, e.g. restenosis, to limit platelet adhesion during interventional procedures or to generate therapies for bone degenerative and neurodegenerative diseases, e.g. Alzheimer's. Agonists may be useful in wound repair.

XX SQ Sequence 1209 BP; 254 A; 330 C; 273 G; 352 T; 0 other;

Query Match 43.9%; Score 240.8; DB 24; Length 1209;
Best Local Similarity 93.7%; Pred. No. 2.3e-55;
Matches 251; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 281 CCGGTCATTCTCTCAGGAACCCCAATGATAAATATGAACCATTTGGGAGGATGAGGA 340
DB 51 CCGGCTGTCTGCCCGCACCCGCGCGGCGATGAATATGAACCATTTGGGAGGATGAGGA 110
QY 341 GAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCT 400
DB 111 GAAAATGAAAGTGGGTTAACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCT 170
QY 401 TCAAAACAACTTCTGCAATTCATCTCAGAGATGCCCTCCGATATTTGACACGCTCTG 460
DB 171 TCAAAACAACTTCTGCAATTCATCTCAGAGATGCCCTCCGATATTTGACACGCTCTG 230
QY 461 GGTGACACTCTTTGTCCTCCATCTGTGTACACCGGAGTGTGTAGTACGCTCCCACTAAA 520
DB 231 GGTGACACTCTTTGTCCTCCATCTGTGTACACCGGAGTGTGTAGTACGCTCCCACTAAA 290
QY 521 CATCATGGCCATCGTTGTGTTCATCTG 548
DB 291 CATCATGGCCATCGTTGTGTTCATCTG 318

RESULT 14
AAT93367
ID AAT93367 standard; cDNA; 1312 BP.
XX AC AAT93367;
XX DT
XX 01-MAY-1998 (first entry)
XX Mouse thrombin receptor cDNA.
XX DE
XX Transgenic animal; thrombin receptor; animal model; restenosis;
KW knockout animal; antagonist; agonist; mouse; murine;
KW Alzheimer's disease; wound repair; ss.
XX OS
XX Mus sp.
XX Location/Qualifiers
FH complement (612..631)
FT primer_bind /*tag= a
FT primer_bind /label= MTR-3
FT primer_bind 1294..1311 /*tag= b
FT primer_bind /label= MTR-2
XX WO9740138-A1.
XX 30-OCT-1997.
XX 22-APR-1997; 97WO-US06575.
XX 25-JUN-1996; 96US-0020544.
XX 23-APR-1996; 96US-0016070.
XX (ORTH) ORTHO PHARM CORP.
XX Andrade-Gordon P, Leung W, Zhou L;
XX WPI; 1997-535823/49.
XX Transgenic animals with altered form of thrombin receptor gene -
PT useful in study of receptor function and identification of
PT antagonists and agonists for disease therapy
XX Example 1; Fig 1; 47pp; English.

KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN W0200200928-A2.
XX
PD 03-JAN-2002.
XX
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
PI
XX WPI; 2002-130909/17.
DR
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation
XX
XX
PS Claim 1; SEQ ID NO 654; 32pp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
XX
SQ Sequence 6436 BP; 1509 A; 261 C; 1477 G; 3189 T; 0 other;

Query Match 27.0%; Score 148.2; DB 24; Length 6436;
Best Local Similarity 72.5%; Pred. No. 5.9e-30;
Matches 192; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1 CGCCGAGGTCGCTGGACCCCTGATCTTACCCGTGGGACCCCTGCTGCTGCTGCCGCG 60
Db 1249 CGCGAAATCGCTTAAACCCCTAATCTTACCCGTAAACACCCCTACCTACCGCG 1190

QY 61 AAGACCGGCTCCCGACCGCGAGAGTCAAGGAGAGGTTGAAGCGAGAGCCGCGGC 120
Db 1189 ARAACCGACTCCCGACCGCGCAAAATCAAAAAAATAAAGAAACAAACCCGAAAC 1130

QY 121 GGGGACGCTCCCGAGCAGCGCGCGCAGAGCCCGGACAATGGGCGCGCGGCTGC 180
Db 1129 GAAACAACCTCCCGAAACAACCGCGCGCAAAACCCGAAACAATAAACCAGGAGCTAC 1070

-QY 181 TGCTGTGGCGGCTGCTTCACTGCTGTGGCGCGCGCTGTGTCGCCGACCCGGGCC 240
Db 1069 TACTAATAACCGGCTTCTTCAATATATACCGCGCTATTATCTACCGCGACCGGAACC 1010

QY 241 GCAGGCGCAGAAATCAAAAGCAACAAA 265
Db 1009 GCAACCAATAAATAATACACGNA 985

Search completed: December 9, 2002, 04:16:37
Job time : 318 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 04:10:21 ; Search time 93 seconds
(without alignments)
2297.668 Million cell updates/sec

Title: US-09-744-679-7
Perfect score: 548
Sequence: 1 cgcgaggtcgttgacc.....ccatcgtgtgttcactcgtg 548

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44.6	8.1	192	9	US-09-925-664-33
2	40.6	7.4	189	9	US-09-925-664-31
3	40.2	7.3	32195	10	US-09-764-870-611
4	40.2	7.3	32195	10	US-09-764-870-617
5	40.2	7.3	32195	10	US-09-764-869-1605
6	39.6	7.2	409	10	US-09-960-352-13321
7	37.8	6.9	286	9	US-10-096-549-4
8	37.6	6.9	2850	12	US-10-074-527-1
9	37.6	6.9	14486	9	US-09-764-868-1504
10	37.4	6.8	1631	10	US-09-925-300-528
11	37.4	6.8	43058	10	US-09-954-456-292
12	37.4	6.8	43058	10	US-09-954-456-529
13	37.4	6.8	43058	10	US-09-880-107-3950
14	37.2	6.8	193	10	US-09-864-761-21978
15	37.2	6.8	12425	12	US-10-023-529-50
16	37.2	6.8	12425	12	US-10-023-523-50
17	37	6.8	80959	9	US-09-858-546-3
18	36.4	6.6	1407	10	US-09-833-381-1620
19	36	6.6	3715	10	US-09-880-107-2300

```

20 35.8 6.5 333 10 US-09-960-352-5606 Sequence 5606, Ap
c 21 35.8 6.5 2561 12 US-10-023-529-48 Sequence 48, Appl
c 22 35.8 6.5 2561 12 US-10-023-523-48 Sequence 48, Appl
23 35.6 6.5 1746 12 US-10-074-527-3 Sequence 3, Appl
c 24 35.4 6.5 930 9 US-09-960-631A-4 Sequence 4, Appl
c 25 35.2 6.4 4973 12 US-10-044-090-293 Sequence 233, App
c 26 35.2 6.4 5119 9 US-09-887-540A-1 Sequence 1, Appl
c 27 35 6.4 2538 9 US-10-136-573A-1 Sequence 1, Appl
c 28 35 6.4 2538 9 US-09-877-665-1 Sequence 1, Appl
c 29 35 6.4 2538 9 US-09-817-647-1 Sequence 1, Appl
c 30 34.8 6.4 1224 10 US-09-815-242-7867 Sequence 7667, Ap
c 31 34.8 6.4 2175 10 US-09-834-975-795 Sequence 795, App
c 32 34.8 6.4 88421 9 US-09-976-059-1 Sequence 1, Appl
c 33 34.6 6.3 405 10 US-09-864-761-6167 Sequence 6167, Ap
c 34 34.6 6.3 1110 10 US-09-893-737-49 Sequence 49, Appl
c 35 34.6 6.3 2750 10 US-09-905-983-51 Sequence 51, Appl
c 36 34.6 6.3 4000 10 US-09-954-456-2153 Sequence 2153, Ap
c 37 34.6 6.3 4000 10 US-09-880-107-3803 Sequence 3803, Ap
c 38 34.6 6.3 42999 9 US-09-772-304A-1 Sequence 1, Appl
c 39 34.6 6.3 42999 9 US-09-799-462A-17 Sequence 17, Appl
c 40 34.6 6.3 42999 9 US-10-125-767-17 Sequence 17, Appl
c 41 34.4 6.3 916 10 US-09-764-870-123 Sequence 123, App
c 42 34.4 6.3 1668 10 US-09-993-811-13 Sequence 13, Appl
c 43 34.4 6.3 2499 10 US-09-350-259-96 Sequence 96, Appl
c 44 34.4 6.3 3124 9 US-09-974-298-128 Sequence 128, App
c 45 34.4 6.3 3230 12 US-10-002-600-47 Sequence 47, Appl

```

ALIGNMENTS

```

RESULT 1
US-09-925-664-33
; Sequence 33, Application US/09925664
; Patent No. US20020160006A1
; GENERAL INFORMATION:
; APPLICANT: Denney, Jr., Dan W.
; TITLE OF INVENTION: Methods of Treating Lymphoma and Leukemia
; FILE REFERENCE: GENITOP-06499
; CURRENT APPLICATION NUMBER: US/09/925,664
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/370,453
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: 08/644,664
; PRIOR FILING DATE: 1996-05-01
; PRIOR APPLICATION NUMBER: 08/761,277
; PRIOR FILING DATE: 1996-12-06
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 192
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(189)
; OTHER INFORMATION:
US-09-925-664-33

```

Query Match 8.1%; Score 44.6; DB 9; Length 192;
Best Local Similarity 71.1%; Pred. No. 0.00093;
Matches 59; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

QY 273 TTAGATCCCGGTCATTTCTTCAGAACCCCAATGATAAATATGACCAATTTGGGAG 332
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
DB 1 TTGGATCCACGATCGTTCTTATTGCGCAATCCAAATGATAAGTACGAACCAATTTGGGAA 60

```

```

QY 333 GATGAGGAGAAAATGAAAGTGG 355
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 61 GATCAGAGCAAGATGCTGAGTGG 83

```



```

CORRESPONDENCE ADDRESS:
  ADDRESSEE: Townsend and Townsend and Crew LLP
  STREET: Two Embarcadero Center, Eighth Floor
  CITY: San Francisco
  STATE: California
  COUNTRY: USA
  ZIP: 94111-3834
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/10/096,549
  FILING DATE: 11-Mar-2002
  CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US/09/242,244
  FILING DATE: 29-Feb-2000
  APPLICATION NUMBER: WO PCT/US97/15270
  FILING DATE: 14-AUG-1997
  APPLICATION NUMBER: US 60/024,213
  FILING DATE: 15-AUG-1996
  ATTORNEY/AGENT INFORMATION:
    NAME: Weber, Kenneth A.
    REGISTRATION NUMBER: 31,677
    REFERENCE/DOCKET NUMBER: 15280-2641PC
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (415) 576-0200
    TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 4:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 286 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    FEATURE:
      NAME/KEY: -
      LOCATION: 1..286
      OTHER INFORMATION: /note="human heat shock protein
        (hsp70b) promoter"
  SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-096-549-4

Query Match          6.9%;      Score 37.8;  DB 9;  Length 286;
Best Local Similarity 60.0%;      Pred. No. 0.11;
Matches 63;  Conservative 0;  Mismatches 42;  Indels 0;  Gaps

QY  69  CTCCCGGACCGCGAGAGTCAGGAGAGGGTGAAGCGGAGCAGCCCGGCGGCGAGC 128
      |||||
Db    227  CTTCCCGACGCTTCCCGCCAGTCGAGAGCCTCCGCGCCGCCGCGGCTGAGTCAGC 168
      |||||

QY  129  CTCCCGGAGCAGCGCCGCGAGAGCCGCGGACAAATGGGCGCGGG 173
      | ||| ||||| ||||| ||||| ||| |||||
Db    167  CGGCGCGGCGAGCGCGCGCTGACCCCGGGAGAGGTTGCGGCGG 123
      | ||| ||||| ||||| ||||| ||| |||||

RESULT 8
US-10-074-527-1
  Sequence 1, Application US/10074527
  Patent No. US20020142426A1
  GENERAL INFORMATION:
    APPLICANT: Olandt, Peter J.
    APPLICANT: Meyers, Rachel E.
    APPLICANT: Galvin, Katherine A.
    APPLICANT: Millennium Pharmaceuticals Inc.
    TITLE OF INVENTION: 33945, A Human Glycosyltransferase and
    TITLE OF INVENTION: Uses Therefor
    FILE REFERENCE: MPI2001-018P(RCPI(M)
    CURRENT APPLICATION NUMBER: US/10/074,527
    CURRENT FILING DATE: 2002-02-12
    PRIOR APPLICATION NUMBER: 60/269202

```

```
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2850
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81)...(1826)
US-10-074-527-1
Query Match      6.9%; Score 37.6; DB 12; Length 2850;
Best Local Similarity 51.8%; Pred. No. 0.42;
Matches 85; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY    90 GGAGAGGGTCAACGCGAGCAGCCCGGAGCGGGGAGCGCTCCCGAGCAGCGCGCGCA 149
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    72 GCGGCGGCATGTGGGCGGCACCGCGCGGGCGCTGCCCGGGAATGCGGCGCGC 131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY    150 GAGCCCCGGACAAATGGGCGCGCGCTGCTGTGTGGCGCCCTCTTCAGTTGTGTC 209
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    132 CGGAGGCGCTTGTGCTCTCTTGCGCTACTGGCGTTGGCGGCTGGCTCGGTGCTG 191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY    210 GGCCCCGCTGTTGTGCGCCGACCGCGCGCGCGAGGCGCAAAATC 253
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB    192 CGGGCGCAGCTGGGCGCGGGCGCGGCTGCGAGCGGGACC 235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 9
US-09-764-868-1504/c
; Sequence 1504, Application US/0974868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1504
; LENGTH: 14486
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-1504
Query Match      6.9%; Score 37.6; DB 9; Length 14486;
Best Local Similarity 53.8%; Pred. No. 1;
Matches 99; Conservative 0; Mismatches 84; Indels 1; Gaps 1;
QY    66 CGGCTCCCAGCCCGCAGAAAGTCAGAGAGAGGGTGAAGCGGAGCAGCCCGCGGGGC 125
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
DB    4230 CGGGCCCCCAACTGACGAGGAGGACG-TAGAGCGGGGCGTGAGCGCCCGGAGCCTCGCT 4172
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
QY    126 AGCTCCCGAGCAGCGCGCAGAGCCGGGACAATGGGCGCGCGCGCTGCTGCTG 185
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
DB    4171 CGGCTCCCGGAGCCCTCCCTCCCGAGCCCGCGCTCACTTGAGCTCCCGGCGCTGAGG 4112
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
QY    186 GTGGCCGCTGCTTCACTAGCTGTGTGCGCGCGCTGTTGTCTGCCCGCACCCGCGCGCAGG 245
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
DB    4111 CGCAGCAGCGCTTGGACACCAAGCTCCGACTCCTCTTCTGCTCTCCCGCGCTCAGC 4052
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
QY    246 CCAG 249
      |||
DB    4051 TCCG 4048
      |||
RESULT 10
US-09-925-300-528
; Sequence 528, Application US/09925300
; APPLICANT: Craig Rosen,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 528
; LENGTH: 1631
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1628)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-300-528
Query Match      6.8%; Score 37.4; DB 10; Length
Best Local Similarity 60.2%; Pred. No. 0.36;
Matches 62; Conservative 0; Mismatches 41; Indels
QY    118 GCGGGGCGAGCTCCGGAGGAGCGCGCGAGAGCGCGGACAAATGGCGC
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||
DB    39 GCGCGGGAGCGCACCCGAGCGAGCGAGCGAGCGCGAGCGCGGAGCGCGGCGGCG
      | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
QY    178 TGCTGTGTGTGGCGCCCTTCAGTCTGTGCGGCGCGCTGTT 220
      | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
DB    99 CGCGCGCGGAGCAGCAACCATGTCTGTTTCGGAAGCTGTT 141
      | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
RESULT 11
US-09-954-456-292
; Sequence 292, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Th
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 292
; LENGTH: 43058
; TYPE: DNA
; ORGANISM: Homo sapiens
```


US-09-954-456-292

Query Match	6.8%	Score 37.4;	DB 10;	Length 43058;
Best Local Similarity	51.7%;	Pred. No. 2;	Mismatches	0;
Matches 109;	Conservative	Indels	101;	Gaps 1;
Qy 11	CGCTTTGACCTGATCTTACCCTGGGACACCTCGGCTCTGCCTCCGCGAGACCGGCT 70			
Db 25326	CGCTGGACCCGGTCACTTCAAGTGGCGGGGCGCGGTGCGGGCGGGCGGGCGGGG 25385			
Qy 71	CCCGACCCCGAGAAAGTCAAGAGAGAGGGTGAACGGAGCAGCCCGAGCGGGGCGAGCCT 130			
Db 25586	CCCGGGGCGGGCGGGGCGCGGGCGGGG-GTCGCGGGCGGGCGGGTGGGTGCGGG 25644			
Qy 131	CCCGGAGCAGCGCGCGCAGACCCCGGACAAATGGGCGCGGGCGGCTGCTGCTGTGTGGC 190			
Db 25645	GGCGGGCGGGTTCGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGGCC 25704			
Qy 191	CGCCTGCTTCAAGTGTGTGCGGGCCCGCTGTGTG 221			
Db 25705	CGCGGGGCGGGGCGGGGCGGGGCGGGGCGGGAGGGGCTG 25735			

RESULT 12
US-09-954-456-529
; Sequence 529, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456

```

: CURRENT APPLICATION NUMBER: US/09/954,456
: CURRENT FILING DATE: 2001-09-18
: PRIOR APPLICATION NUMBER: US/60/233,617
: PRIOR FILING DATE: 2000-09-18
: PRIOR APPLICATION NUMBER: US/60/234,052
: PRIOR FILING DATE: 2000-09-20
: PRIOR APPLICATION NUMBER: US/60/234,923
: PRIOR FILING DATE: 2000-09-25
: PRIOR APPLICATION NUMBER: US/60/235,134
: PRIOR FILING DATE: 2000-09-25
: PRIOR APPLICATION NUMBER: US/60/235,637
: PRIOR FILING DATE: 2000-09-26
: PRIOR APPLICATION NUMBER: US/60/235,638
: PRIOR FILING DATE: 2000-09-26
: PRIOR APPLICATION NUMBER: US/60/235,711
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: US/60/235,720
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: US/60/235,840
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: US/60/235,863
: PRIOR FILING DATE: 2000-09-27
: NUMBER OF SEQ ID NOS: 2276
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 529
: LENGTH: 43058
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-954-456-529

```

Query Match	6.8%	Score 37.4	DB 10	Length 43058
Best Local Similarity	51.7%	Pred. No. 2		
Matches 109	Conservative 0	Mismatches 101	Indels 1	Gaps 1
11	CGCTTGACCCCTGATCTTACCCCTGGGCGACCCCTGCGCTCTGCGTCCGCGAGACCGGCT	70		
2526	CGCGTGACCCCGCTCAACTTCAAGGTGCGGGGGCGGTCGCGGGCGGGCGGGCGGGG	25585		
71	CCCCGACCCCGACAGCTCAGGAGAGAGGGTGAAGCGGAGCACCCCGAGCGGGCGGCGAGCCT	130		
25586	CCGCGGGCGGGCGGGCGGGCGCGGGGGCGGGG-GTCTCGGGGCGGGGCGGGGTCGGGTCGGGG	25644		

	QY	131	CCGGAGACGCCCGCCGACAGACC CGGCATATGGGGCGCGGGCTGTCTGTTGTC	190
	Db	25645	GCGGGGCGGGGTCCGCGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGC	25704
	QY	191	C GCCTGCTTCA G T C T G T C G C G C C C G C T G T T G	221
	Db	25705	C GGCGGGCGGGGCGGGGCGGGCGGGGAGGGCTG	25735

```

RESULT 13
US-09-880-107-3950
; Sequence 3950, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3950
; LENGTH: 43058
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; - OTHER INFORMATION: Genbank Accession No. US20020142981A1 284721
US-09-880-107-3950

```

	Query Match	6.8%	Score 37.4;	DB 10;	Length 43058;
	Best Local similarity	51.7%;	Pred. No. 2;	Mismatches 101;	Indels 1; Gaps 1;
	Matches 109;	Conservative 0;			
QY	11	CGTWTGGACCCCTGATCTTACCCTGTGGGCACCTCGCTCTGCCCTGCCGCGAGACCGGCT	70		
Db	25526	C GG GTTGACCCCGTCAACTTCAGGTGCGGGGCGGGTGTGCGGCGGGCGGGCGGGG			
QY	71	CCCCGACCCCGCAGAAGTCAGGAGAGGGTGAACGCGAGCACGCCAGAGGGGGGCGACGCT	130		
Db	25586	CGCGGGGGCGGGGGGGCGCGGGGGGG -GTCCGGGGCGGGCGGGTGGGTCTCGGG	25644		
QY	131	CCCGGAGCAGCGCGCGCAGAGCCC CGGACAAATGGGCCCGCGCGGCTGTCTGTGGTGGC	190		
Db	25645	GCGGGGCGGGGTCTCGGGGCGGGGCGGGCGGGCGGGCGGGCGGGCGGGGCC	25704		
QY	191	CGCCTGCTTCAAGTCTGTGGCGGCCCGCTGTG	221		
Db	25705	CGCGGGGGCGGGCGGGGGCGGGGGGAGGGGCTG	25735		

```

RESULT 14
US-09-864-761-21978
; Sequence 21978, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 04:06:36 ; Search time 2427 Seconds
(without alignments)
3656.830 Million cell updates/sec

Title: US-09-744-679-7

Perfect score: 548

Sequence: 1 cgcgcagggtcgcttgacc.....ccatggtgtgttcattcctg 548

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *

1: em_estba:*

2: em_esthum:*

3: em_estmu:*

4: em_estm:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	548	100.0	888	14	BQ440574
2	536	97.8	699	10	BE258615
3	505.2	92.2	739	12	BG473471
4	197	35.9	496	11	BC016059
5	183	33.4	621	10	BB662299
6	183	33.4	674	13	BI685426

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	183	33.4	942	14	BQ896389
8	182.4	33.3	949	14	BQ923464
9	180.8	33.0	584	13	BI738768
10	175.8	32.1	688	9	AA882290
11	165.6	30.2	672	12	BE911999
12	162.2	29.6	763	13	BI666195
13	159.2	29.1	467	10	BB862683
14	157.6	28.8	466	10	BB852109
15	152	27.7	1429	14	BQ926183
16	145.8	26.6	405	12	BF116952
17	133	24.3	557	12	BQ667569
18	132	24.1	554	14	C87335
19	132	24.1	588	14	C86764
20	128	23.4	504	9	AA690768
21	126.2	23.0	486	10	BE128304
22	126	23.0	350	12	BF407166
23	124.6	22.7	475	9	AA096771
24	122.2	22.3	462	12	BG744398
25	118.2	21.6	906	13	BI729686
26	115	21.0	436	10	BB864321
27	94.8	17.3	370	10	BB866629
28	90.6	16.5	316	10	BE126861
29	85.8	15.7	273	12	BB864515
30	84	15.3	375	10	BE127249
31	72.4	13.2	272	10	BB838066
32	70.2	12.8	326	10	BB865249
33	67.8	12.4	683	9	AJ449531
34	67.8	12.4	704	9	AJ449546
35	67.8	12.4	824	9	AJ456565
36	67.4	12.3	718	9	AJ447241
37	66.2	12.1	798	9	AJ455565
38	57.2	10.4	127	14	W20934
39	51.6	9.4	925	17	CNS0091P
40	51	9.3	910	17	CNS0060N
41	51	9.3	925	17	CNS0091P
42	50.4	9.2	935	17	CNS006XK
43	49.6	9.1	1212	14	BQ669498
44	48.4	8.8	932	17	CNS0072Q
45	48.2	8.8	1010	17	CNS016TE

ALIGNMENTS

RESULT 1	BQ440574	AGENCOURT_7907353	NIH_MGC_67	Homo sapiens	cdna	linear	EST 24-MAY-2002
BQ440574	LOCUS	5', mRNA sequence.	888 bp	mrna	linear	EST 24-MAY-2002	6154861
DEFINITION	AGENCOURT_7907353	NIH_MGC_67	Homo sapiens	cdna	linear	EST 24-MAY-2002	6154861
ACCESSION	BQ440574	5', mRNA sequence.					
VERSION	BQ440574.1	GI:21179650					
KEYWORDS	EST.						
SOURCE	human.						
ORGANISM	Homo sapiens						
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.						
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.						
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)						
JOURNAL	Unpublished (1999)						
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1AM13496 row: j column: 14 High quality sequence stop: 654. Location/Qualifiers 1. .888						

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6154861"
/clone_lib="NIH_MGC_67"
/tissue_type="retinoblastoma"
/lab_host="Phi10B (phage-resistant)"
/notes="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dm.
Average insert size 1.75 kb. Library constructed by Life
technologies."

```

BASE COUNT	181 a	263 c	237 g	207 t	
ORIGIN					
Query Match	100.0%	Score 548;	DB 14;	Length 888;	
Best Local Similarity	100.0%;	Pred. No. 1.1e-127;			
Matches 548;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy 1	CGCCGAGGTCGCTTGACCCCTGATCTTACCCGTGGGACACCTCGCTCTCCCTGCCGCG	60			
Db 25	CGCCGAGGGTGCCTTGGACCTGATCTTACCCGTGGGACACCTCGCTCTCCCTGCCGCG	84			
Qy 61	AAGACCGGCTCCCCGACCCGAGAGTTCAGGAGAGAGGGTGAAGGGGAGAGCCCGAGGC	120			
Db 85	AAGACCGGCTCCCCGACCCGAGAGTTCAGGAGAGAGGGTGAAGGGGAGAGCCCGAGGC	144			
Qy 121	GGGCGAGCCTCCCGGAGCAGCGCGCAGAGCCGGGACAATGGGGCGCGGGCGGCTGC	180			
Db 145	GGGCGAGCCTCCCGGAGCAGCGCGCGCAGAGCCGGGACAATGGGGCGCGGGCGGCTGC	204			
Qy 181	TGCTGGTGGCGCCTGCTTCAGTCTGTGTGGGCGCGCTGTGTCTCCCGCACCCGGGCCC	240			
Db 205	TGCTGGTGGCGCCTGCTTCAGTCTGTGTGGGCGCGCTGTGTCTCCCGCACCCGGGCCC	264			
Qy 241	CGAGCCGAGAAATCAAAAGCAACAAATGCACCTTAGATCCCGGTCATTTCTTCTCAGGA	300			
Db 265	CGAGCCGAGAAATCAAAAGCAACAAATGCACCTTAGATCCCGGTCATTTCTTCTCAGGA	324			
Qy 301	ACCCCAATGATAAATATGAACCATTTTGGAGGATGAGAGAAAAATGAAGTGGGTAA	360			
Db 325	ACCCCAATGATAAATATGAACCATTTTGGAGGATGAGAGAAAAATGAAGTGGGTAA	384			
Qy 361	CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCCTCTTCAAAAACAACCTTCCTGCAT	420			
Db 385	CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCCTCTTCAAAAACAACCTTCCTGCAT	444			
Qy 421	TCATCTCAGAAGATGCCCTCCGGATATTGACGAGCTCTGGGCTGACACTCTTTGTCCCAT	480			
Db 445	TCATCTCAGAAGATGCCCTCCGGATATTGACGAGCTCTGGGCTGACACTCTTTGTCCCAT	504			
Qy 481	CTGTGTACACCGAGTGTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT	540			
Db 505	CTGTGTACACCGAGTGTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT	564			
Qy 541	TCATCCTG 548				
Db 565	TCATCCTG 572				

RESULT 2	
BE258615	
LOCUS	699 bp mRNA linear EST 13-JUL-2000
DEFINITION	G01106874F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:3343051 5', mRNA sequence.
ACCESSION	BE258615
VERSION	BE258615.1
KEYWORDS	GI:9129107
SOURCE	EST.
ORGANISM	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 699)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLC117 row: c column: 20.
Location/Qualifiers
1. .699
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3343051"
/clone_lib="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pDB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCACAGAG(G). Library constructed by Ling Hon
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)
Note: this is a NIH_MGC Library."
153 a 202 g 194 g 149 t 1 others

BASE COUNT	153 a	202 c	194 g	149 t	1 others
ORIGIN					
Query Match	97.8%;	Score 536;	DB 10;	Length 699;	
Best Local Similarity	99.6%;	Pred. No. 1.1e-124;			
Matches 547; Conservative	0;	Mismatches 1;	Indels 1;	Gaps	
Qy	1	CGCGAGGGTCGCTTGGACCTGATCTTACCGTGGCGACCCCTGCCTGTGCCTGCGCGG 60			
Db	51	CGCGAGGGTCGCTTGGACCTGATCTTACCGTGGCGACCCCTGCCTGTGCCTGCGCGG 110			
Qy	61	AAGACCGGCTCCCGACCCGCGAGAGTCAGAGAGAGGGTGAACGCGAGCAGAGCCC -GAG 119			
Db	111	AAGACCGGCTCCCGACCCGCGAGATCAGGAGAGAGGGTGAAGCGGAGCAGAGCCCTGAGG 170			
Qy	120	CGGGCGAGCTCCCGGAGCAGCGCCGCGCAGAGCCGGGCAATGGGGCCCGCGCGCGCTG 179			
Db	171	CGGGCGAGCTCCCGGAGCAGCGCCGCGCAGAGCCGGGCAATGGGGCCCGCGCGCGCTG 230			
Qy	180	CTGCTGTGTGGCGCGCTGCTTTCAGTCTGTGCGGCCCGCTGTGTCTGTGCCGACCCCGGGCC 239			
Db	231	CTGCTGTGTGGCGCGCTGCTTTCAGTCTGTGCGGCCCGCTGTGTCTGTGCCGACCCCGGGCC 290			
Qy	240	CGGAGGCCAGAACTCAAAGCAACAATGCCACTTAGATCCCGGCTCATTTCTTCTCAGG 299			
Db	291	CGGAGGCCAGAACTCAAAGCAACAATGCCACTTAGATCCCGGCTCATTTCTTCTCAGG 350			
Qy	300	AACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAATGAAGTGGCTTA 359			
Db	351	AACCCCAATGATAAATATGAACCATTTTGGGAGGATGAGGAGAAAATGAAGTGGCTTA 410			
Qy	360	ACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTTCAAAAAACAATCTCCCTGCA 419			
Db	411	ACTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTTCAAAAAACAATCTCCCTGCA 470			
Qy	420	TTCATCTCAGAAGATGCGCTCGGATATTGGACCAAGTCTCGCTCAGACTCTTTGTGCCCA 479			
Db	471	TTCATCTCAGAAGATGCGCTCGGATATTGGACCAAGTCTCGCTCAGACTCTTTGTGCCCA 530			
Qy	480	TCGTGTGTACCGGAGTGTGTGTAGTCAGCGCTCCCACTAAACATCATGSCCATGCTTGTG 539			
Db	531	TCGTGTGTACCGGAGTGTGTGTAGTCAGCGCTCCCACTAAACATCATGSCCATGCTTGTG 590			
Qy	540	TTTCATCCCTG 548			
Db	591	TTTCATCCCTG 599			

```

RESULT 3
BG473471
LOCUS
DEFINITION
  BG473471 739 bp mRNA linear EST 21-MAR-2001
  60251627F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4647883 5',
  mRNA sequence.
ACCESSION
  BG473471
VERSION
  BG473471.1 GI:13405746
KEYWORDS
  EST.
SOURCE
  Human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 739)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Ling Hong/Rubin Laboratory
  DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
  Plate: L1CM1425 row: c column: 20
  High quality sequence stop: 718.
FEATURES
  source
    1..739
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4647883"
    /clone_lib="NIH_MGC_16"
    /tissue_type="retinoblastoma"
    /lab_host="DH10B (phage-resistant)"
    /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
    EcoRI; cDNA made by oligo-dT priming. Directionally
    cloned into EcoRI/XhoI sites using the following 5'
    adaptor: GGCACGAG(G). Library constructed by Ling Hong
    in the laboratory of Gerald M. Rubin (University of
    California, Berkeley) using ZAP-cDNA synthesis kit
    (Stratagene) and Superscript II RT (Life Technologies).
    Note: this is a NIH_MGC library."
  BASE COUNT 162 a 207 c 200 g 170 t
  ORIGIN

Query Match 92.2%; Score 505.2; DB 12; Length 739;
Best Local Similarity 99.1%; Pred. No. 6.6e-117;
Matches 529; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 16 GCACCTGATCTACCGTGGGACCCCTGCGTCTGCTGCGCGAAGACCGGCTCCCG 75
|
|
|
Db 2 GGACCTGATCTTACCGTGGGACCCCTGCGTCTGCTGCGCGAAGACCGGCTCCCG 61
|
|
|
QY 76 ACCCGCAGAGTCAGGAGAGGGTGAAGCGGAGCAGCCCGGAGGCGGCGAGCCTCCCG 135
|
|
|
Db 62 ACCCGCAGAGTCAGGAGAGGGTGAAGCGGAGCAGCCCGGAGGCGGCGAGCCTCCCG 121
|
|
|
QY 136 AGCAGCGCGCGCAGAGCGCGGACAAATGGGCGCGCGGCGTGTCTGTGTGGCGGCGCT 195
|
|
|
Db 122 AGCAGCGCGCGCAGAGCGCGGACAAATGGGCGCGCGGCGTGTCTGTGTGGCGGCGCT 181
|
|
|
QY 196 GCTTCAGTCTGTGGCGCGCGTGTGTCTGTGCGCGCGCGCGCGAGCGCAGAAATCAA 255
|
|
|
Db 182 GCTTCAGTCTGTGGCGCGCGTGTGTGTGCGCGCGCGCGCGCGAGCGCAGAAATCAA 241
|
|
|
QY 256 AAGCAACAAATGCCACCTTAGATCCCGGTCATTTCTCTCAGGAACCCCAATGATAAT 315
|
|
|
Db 242 AAGCAACAAATGCCACCTTAGATCCCGGTCATTTCTCTCAGGAACCCCAATGATAAT 301
|
|
|
QY 316 ATGAACCAATTTGGGAGGATGAGGAGAAAAATGAAGTGGGTAACTGAATACAGATTAG 375
|
|
|
Db 302 ATGAACCAATTTGGGAGGATGAGGAGAAAAATGAAGTGGGTAACTGAATACAGATTAG 361
|
|
|

```

```

QY 376 TCTCCATCAATAAAGCAGTCTCTTCAAAAAACAACCTCC-TGCATTCTCAGAGAT 434
|
|
|
Db 362 TCTCCATCAGTAAGACGAGTCTCTTCAAAAAACAACCTCTTGCATTCATCAGAGAT 421
|
|
|
QY 435 GCCTCCGGATATTTGACCAAGCTCCTGGGTGACACACTCTTTGTCCTCTGTACACCGGA 494
|
|
|
Db 422 GCCTCCGGATATTTGACCAAGCTCCTGGGTGACACACTCTTTGTCCTCTGTGTACACCGGA 481
|
|
|
QY 495 GTGTTTGTAGTCAGCTCCCACTAAACATCATGGCCATCGTTGTCTTCATCTG 548
|
|
|
Db 482 GTG-TTGATGTCAGCTCCCACTAAACATCATGGCCATCGTTGTCTTCATCTG 534
|
|
|
RESULT 4
BC016059
LOCUS
DEFINITION
  BC016059 496 bp mRNA linear HTC 24-OCT-2001
  Clone IMAGE:4849569, mRNA.
ACCESSION
  BC016059
VERSION
  BC016059.1 GI:16359189
KEYWORDS
  HTC.
SOURCE
  Homo sapiens.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 496)
  Strausberg, R.
  Direct Submission
  Submitted (22-OCT-2001) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
  Email: cgapbs-r@mail.nih.gov
  Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
  cDNA Library Preparation: Rubin Laboratory
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Genome Sequence Centre,
  BC Cancer Agency, Vancouver, BC, Canada
  info@bcgsc.bc.ca
  Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
  Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
  Letitia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
  Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
  Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
  Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
  Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
  George Yang, Scott Zuyderduyn, Marco Marra.
  Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
  Series: IRAL Plate: 34 Row: a Column: 7
  This clone was selected for full length sequencing because it
  passed the following selection criteria: matched mRNA gi: 6031164
  This clone has the following problem: incomplete processing.
FEATURES
  source
    1..496
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:4849569"
    /tissue_type="Blood, natural killer cell"
    /clone_lib="NIH_MGC_106"
    /lab_host="DH10B-R"
    /note="Vector: pOTB7"
  BASE COUNT 130 a 144 c 146 g 76 t
  ORIGIN

```

```

Query Match 35.9%; Score 197; DB 11; Length 496;
Best Local Similarity 100.0%; Pred. No. 3.3e-39;
Matches 197; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	53	CTGCCGGAACCGGTCCCGCAGCAAGTTCAGGAGAGGGTGAAAGCGGAGCAC	112
Dd	1	CTCCGCGAAGACCGGTCCCGCAGCAAGTTCAGGAGAGGGTGAAAGCGGAGCAC	60
QY	113	CCGAGCGGGGAGCGTCCCGGAGCAGCCCGCGCAGAGCCCGGACAAATGGGGCGCG	172
Dd	61	CCGAGCGGGGAGCGTCCCGGAGCAGCCCGCGCAGAGCCCGGACAAATGGGGCGCG	120
QY	173	GCGGCTCTGTTGGTGCCTGCTTCACGTCTGTGGCGCCCGCTGTGTCTCTCCGCGCAC	232
Dd	121	GCGGCTCTGTTGGTGCCTGCTTCACGTCTGTGGCGCCCGCTGTGTCTCTCCGCGCAC	180
QY	233	CCGGCCCCCGCAGGCGCAG	249
Dd	181	CCGGCCCCCGCAGGCGCAG	197

RESULT 5	BB662299	621 bp	mRNA	linear	EST 26-OCT-2001
LOCUS	BB662299	RIKEN full-length enriched, 16 days neonate heart Mus			
DEFINITION	Musculus CDNA clone DB30039D11 5', mRNA sequence.				
ACCESSION	BB662299				
VERSION	BB662299.1	GI:16496054			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
	1 (bases 1 to 621)				
	Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,				
	Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,				
	M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,				
	Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,				
	D., Shibata,K., Shinagawa,A., Shiraki,T., Sugabe,Y., Suzuki,H.,				
	Tagam,i,M., Tagawa,M., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,				
	Muramatsu,M. and Hayashizaki.Y.				
	RIKEN Mouse ESTs (Arakawa,T., et al. 2001)				
TITLE	Unpublished (2001)				
JOURNAL	Contact: Yoshhide Hayashizaki				
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic				
	Sciences Center(GSC), Yokohama Institute				
	The Institute of Physical and Chemical Research (RIKEN)				
	1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan				
	Tel.: 81-45-503-9222				
	Fax: 81-45-503-9216				
	Email: genome-res@gsc.riken.go.jp,				
	URL:http://genome.gsc.riken.go.jp/				
	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh				
	M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki.Y.				
	Normalization and subtraction of cap-trapper-selected cDNAs to				
	prepare full-length cDNA libraries for rapid discovery of new				
	genes. Genome Res. 10 (10), 1617-1630 (2000)				
	wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,				
	Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura				
	S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and				
	Hayashizaki,Y.				
	RIKEN integrated sequence analysis (RISA) system--384-format				
	sequencing pipeline with 384 multicapillary sequencer. Genome Res.				
	10 (11), 1757-1771 (2000)				
	Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,				
	I., and Hayashizaki,Y.				
	Computer-based methods for the mouse full-length cDNA				
	encyclopedia: real-time sequence clustering for construction of a				
	non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001)				
	Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,				
	K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and				
	Hayashizaki,Y.				
	Computational Analysis of Full-Length Mouse cDNAs Compared with				
	Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)				
	Please visit our web site (http://genome.gsc.riken.go.jp) for				
	further details.				
	e mouse tissues.				

Db 384 CTGACGCTCTTCATGCGCCCTCCGCTGTACACAGATTGTTCATTCTAGCCTTCTCTGAAC 443

QY 522 ATCATGCCATCGTGTGTTTCATCCCTG 548

Db 444 GTCTGGCCATCGACGTTGTCGTCCTG 470

RESULT 8

LOCUS BQ923464

DEFINITION AGENCOURT_8800954 NIH_MGC_130 Mus musculus cDNA clone IMAGE:6392393

ACCESSION BQ923464

VERSION BQ923464.1 GI:22338495

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 949)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman, Ph.D.
cDNA Library Preparation: ResGen, Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13882 row: c column: 18
High quality sequence stop: 674.

FEATURES

source 1..949

Location/Qualifiers

1..949

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:6392393"

/clone_lib="NIH_MGC_130"

/lab_host="DH10B (phage-resistant)"

/note="Organ: otocysts; Vector: pCMV-SPORT6.1.ccdB;
Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.
Primer: Oligo dT. Average insert size 1.95 kb.
Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."

BASE COUNT 161 a 315 c 261 g 212 t

ORIGIN

Query Match 33.3%; Score 182.4; DB 14; Length 949;

Best Local Similarity 67.9%; Pred. No. 1.9e-35;

Matches 277; Conservative 0; Mismatches 116; Indels 15; Gaps 1;

QY 156 GGGACAATGGGCGCGGGCGTGTGTGTGGCGGCTTCAGTCTGTGGGCGCG 215

Db 41 GGGACAATGGGCGCGGGCGTGTGTGTGTGGCGGCTTCAGTCTGTGGGCGCG 100

QY 216 CTGTGTCTGCGCGACCGCGCGCGCGAGATCAAAAGCAAAATGCCACCTTA 275

Db 101 TTGCTGTCTTCCCGCGTCTATGAGCCAGCCAGAGATCAGAGAGACAGATGCTACGGTG 160

QY 276 GATCCCGGTCTATTTCTCTCAGGAACCCCAATGATAATATGAACCATTT----- 325

Db 161 ACCCCCGCTCATTTCTTCTAAGGATCCCACTGAAATATACATTTGAATGGTCCCGCTG 220

QY 326 -----TTGGGAGATGAGAGAAAATGAAGTGGTGGTGAATGATACAGATTAGTCTCC 380

Db 221 GGGGATGAGGAGGAGGAGGAGAAAATGAAGCGCTGCTGTGGAGGGTAGGGCAGTCTAC 280

QY 381 ATCAATAAAGAGAGTCTCTTCAAAAACAACTTCCTGCAATTCATCTCAGAAGATGCCCTCC 440

Db 281 TTAATATAGAGCTCCTCTCTCACACGGCGCTCTCTCTCTCCCTTCATCTCCGAGGACGCTCC 340

QY 441 GGATATTTGACAGCTCTGCTGACACTCTTTGTCCACTCTGTGTACACGGAGTGT 500

Db 341 GGATATCTGACGACCCCTGGCTGACGCTCTTCATGCCCTCGGTACACGATTGTGTC 400

QY 501 GTAGTCAGCCTCCCACTAAACATCATGCCATCGTGTGTGTTCAATCCCTG 548

Db 401 ATTGTGAGCTTCTCTGAACGCTCTGGCCATCGCAGTGTTCGCTTG 448

RESULT 9

LOCUS B1738768

DEFINITION 603362614F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5369618 5',

ACCESSION B1738768

VERSION B1738768

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 584)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM11940 row: 1 column: 03
High quality sequence stop: 582.

FEATURES

source 1..584

Location/Qualifiers

1..584

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="IMAGE:5369618"

/clone_lib="NIH_MGC_94"

/tissue_type="retina"

/lab_host="DH10B (phage-resistant)"

/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: Sali; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

BASE COUNT 105 a 181 c 168 g 130 t

ORIGIN

Query Match 33.0%; Score 180.8; DB 13; Length 584;

Best Local Similarity 67.6%; Pred. No. 4.2e-35;

Matches 276; Conservative 0; Mismatches 117; Indels 15; Gaps 1;

QY 156 GGGACAATGGGCGCGGGCGTGTGTGTGGCGGCTTCAGTCTGTGGCGCGCG 215

Db 36 GGGACAATGGGCGCGGGCGTGTGTGTGGCGGCTTCAGTCTGTGGCGGCTCC 95

QY 216 CTGTGTCTGCGCGACCGCGCGCGCGAGATCAAAAGCAAAATGCCACCTTA 275

Db 96 TTGCTGTCTTCCCGCTGCTCTATGAGCCAGCCAGAGATCAGAGAGACAGATGCTACGGTG 155

QY 276 GATCCCGGTCTATTTCTTCAGGAACCCCAATGATAATATGAACCATTT----- 325

Db 156 AACCCCGCTCATTTCTTCTAAGGATCCCACTGAAATATACATTTGAACATGGTCCCGCTG 215

QY 326 -----TTGGGAGATGAGGAGAAAATGAAGTGGTGGTAACTGATAACAGATTAGTCTCC 380

Db 216 GGGGATGAGGAGGAGGAGGAGAAAATGAAGCGCTCTCTGTGGAGGGTAGGGCAGTCTAC 275


```

QY 381 ATCAATAAAGCAGCTCTCTTCAAAAACAACTTCTGATTCATCTCAGAGATGCCTCC 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 276 TTAATATAAGCCTCCCTCTCACAGCGCCTCTCTCTCTCTCTCCGAGACGCTCC 335
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 441 GGATATTTGACCAAGTCTCTGGCTGACACTCTTTGTCCTCATCTGTACACCGAGTGT 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 336 GGATATCTGACCAAGCCTCTGGCTGACGCTCTTCATGCTCCTCGGTGTACAGGATGTG 395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 501 GTAGTCAGCTCCCACTAAACATCATGCGCATCTGTGTGTCATCCTG 548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 396 ATTGTACGCTCTCTGAACGCTCTGCGCATGCGAGTGTGCTGTTG 443

RESULT 10
AA882290
LOCUS
DEFINITION
  vx43b01.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone
  IMAGE:1277929 5' similar to gb:M62424 THROMBIN RECEPTOR PRECURSOR
  (HUMAN); mRNA sequence.
ACCESSION
  AA882290
VERSION
  AA882290.1 GI:2991401
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 688)
  Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
  Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
  Schellenberg,K., Stepcoe,M., Tan,F., Underwood,K., Moore,B.,
  Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
  Waterston,R.
  The WashU-HMI Mouse EST Project
  Unpublished (1996)
  Contact: Marra M/Mouse EST Project
  WashU-HMI Mouse EST Project
  Washington University School of MedicineP
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: mouseest@wustl.edu
  This clone is available royalty-free through LLNL ; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:669729
  Seq primer: -28ml3 rev1 ET from Amersham
  High quality sequence stop: 314.
  Location/Qualifiers
    1..688
      /organism="Mus musculus"
      /strain="C57BL/6 x CBA"
      /db_xref="taxon:10090"
      /clone="IMAGE:1277929"
      /clone_lib="Stratagene mouse lung 937302"
      /sex="female"
      /tissue_type="lung"
      /dev_stage="6-8 month old"
      /lab_host="SOLR (kanamycin resistant)"
      ; Site_1: EcoRI
      ; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
      dt. 6-8 month old female lung and 1.5 year old male lung
      were source of mRNA. Average insert size: 1.5 kb; Uni-ZAP
      XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3'
      adaptor sequence: 5' CTCAGATTTTTTTTTTTT 3"
BASE COUNT 126 a 212 c 188 g 162 t
ORIGIN
  Query Match 32.1%; Score 175.8; DB 9; Length 688;
  Best Local Similarity 65.6%; Pred. No. 8e-34;
  Matches 279; Conservative 0; Mismatches 132; Indels 14; Gaps 1;

QY 138 CAGCGCGCGGACAGCCCGGACAAATGGGCGCGCGCTGCTGCTGTGGCGCGCTCG 197
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 CGGCGCGCGGCGCGCTGGGACACATGGGCGCGCGCTTGTGATCGTCCCTCGCG 60

```

```

QY 198 TTCACTCTGTGGGCCCGCTGTGTGTCGCCGACCCGCGCGCAGAGCCAGAAATCAAAA 257
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CTCAGCCTGTGCGATCCCTTGTGTTCTTCCGCGTCCCTATGAGCCAGCCAGATCAGAG 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 258 GCAACAAATGCCACCTTAGATCCCGGTGATTTCTTCTCAGGAACCCCAATGATAAATAT 317
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 AGGACAGATGTACGGTGAACCCCGCTCATTTCTTTCTAAGGAATCCCAAGTAAAAATACA 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 318 GAACCAATT-----TTGGGAGGATGAGGAGAAAAATGAAAGTGGTAACTG 363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 TTTGAACTGGTCCCGCTGGGATGAGGAGGAGAGAGAAAAATGAAAGCTGCTGCTGG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 364 AATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTTCAAAAACAACTTCCTGCATTTCA 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 AGGGTAGGGCAGCTACTTAAATATAAAGCTCCCTCCACACGCGGCTCCTCCCTTCA 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 424 TCTCAGAAGATGCCCTCCGATATTTGACAGCTCTCTGCTGACACTCTTTGTGCCATCTG 483
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 TCTCGAGGACGCTCCCGATATCTGACAGCCCTGGCTGACGCTCTTCATGCGCTCG 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 484 TGTACACCGAGTGTGTGTTAGTACGCTCCCACTAAACATCATGCGCATCGTGTGTTTCA 543
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 TGTACACGATTGTGTTTCTGTTGTCAGCTTCTCTGACGCTCTGCGCATCGCAGTGTTCG 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 544 TCCCTG 548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 421 TCTTG 425
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
BE911999
LOCUS
DEFINITION
  601662765F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3962777 5',
  mRNA sequence.
ACCESSION
  BE911999
VERSION
  BE911999.1 GI:10408985
KEYWORDS
  EST.
SOURCE
  house mouse.
ORGANISM
  Mus musculus
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 672)
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgabs-r@mail.nih.gov
  Tissue Procurement: Gilbert Smith, Ph.D.
  CDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Incyte Genomics, Inc.
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLAM9130 row: a column: 18
  High quality sequence stop: 626.
  Location/Qualifiers
    1..672
      /organism="Mus musculus"
      /strain="FVB/N"
      /db_xref="taxon:10090"
      /clone="IMAGE:3962777"
      /clone_lib="NCI_CGAP_Mam1"
      /tissue_type="tumor, biopsy sample"
      /dev_stage="10 months, virgin"
      /lab_host="DH10B"
      /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
      Site_2: NotI; Cloned unidirectionally. Primer: Oligo df.
      Library constructed by Life Technologies. Investigator
      providing samples: Gilbert Smith, NIH"
BASE COUNT 118 a 211 c 188 g 155 t
ORIGIN

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 9, 2002, 04:05:26 ; Search time 3119 Seconds
(without alignments)
5113.286 Million cell updates/sec

Title: US-09-744-679-7
Perfect score: 548
Sequence: 1 cgcggagggtcgctggacc.....ccatcggtgtgttcacctcg 548

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vt:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vt:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	548	100.0	3472	9	HUMTHRR	M62424 Human throm
2	548	100.0	3479	9	BC002464	BC002464 Homo sapi
3	548	100.0	3480	6	AR025392	AR025392 Sequence
4	548	100.0	3480	6	AR027730	AR027730 Sequence
5	548	100.0	3480	6	AR065265	AR065265 Sequence
6	548	100.0	3480	6	AR137610	AR137610 Sequence
7	548	100.0	3480	6	I74660	I74660 Sequence 21
8	441.6	80.6	3764	9	AF028727	AF028727 Papio ham
9	306.2	55.9	70679	2	AC026151	AC026151 Homo sapi
10	302	55.1	24771	9	AF391809	AF391809 Homo sapi
11	302	55.1	110965	2	AC020927	AC020927 Homo sapi
12	302	55.1	143335	9	AC025188	AC025188 Homo sapi
13	282.4	51.5	1764	6	I49726	I49726 Sequence 12
14	249	45.4	88881	2	AC008933	AC008933 Homo sapi
15	246	44.9	2895	9	HSU63331	HSU63331 Human throm
16	229	41.8	2878	9	AF054633	AF054633 Homo sapi
17	188.2	34.3	2855	10	CLTHRRRC	CLTHRRRC Sequence
18	184.2	33.6	1312	10	MUSTHRRCT	MUSTHRRCT Sequence
19	183	33.4	3313	10	BC031516	BC031516 Mus muscu
20	171.2	31.2	3418	10	RATTRGPC	RATTRGPC Sequence
21	161	29.4	5457	9	HSU36755	HSU36755 Mus muscu
22	148.2	27.0	6436	6	AX345583	AX345583 Sequence
23	146.6	26.8	6436	6	AX345582	AX345582 Sequence
24	137.6	25.1	151166	2	AC130634	AC130634 Rattus no
25	133	24.3	5620	10	MMTHRECO2	MMTHRECO2 Sequence
26	133	24.3	199227	2	AC110876	AC110876 Mus muscu
27	130.4	23.8	1164	10	GLU34047	GLU34047 Cricetulus
28	53	9.7	275	10	MMCF2R01	MMCF2R01 Mus muscu
29	53	9.7	4605	10	MMTHRECO1	MMTHRECO1 Mus muscu
30	53	9.7	235097	2	AC110877	AC110877 Mus muscu
31	52.6	9.6	1680	5	XLU09632	XLU09632 Xenopus lae
32	48.8	8.9	47981	1	AF263245	AF263245 Micromono
33	48.8	8.9	47981	6	AX112026	AX112026 Sequence
34	47.2	8.6	195670	2	AC115449	AC115449 Rattus no
35	47	8.6	2464	10	AY034062	AY034062 Mus muscu
36	47	8.6	69747	2	AC124340	AC124340 Mus muscu
37	47	8.6	193947	10	AL606985	AL606985 Mouse DNA
38	47	8.6	195554	2	AL714009	AL714009 Mus muscu
39	46.4	8.5	86237	2	AC111966	AC111966 Rattus no
40	46	8.4	91856	2	AC123018	AC123018 Rattus no
41	46	8.4	118155	2	AC130148	AC130148 Rattus no
42	45.2	8.2	157861	2	AC126902	AC126902 Rattus no
43	45	8.2	134725	2	AC114389	AC114389 Rattus no
44	45	8.2	175344	2	AC097596	AC097596 Rattus no
45	44.8	8.2	159822	2	AC118304	AC118304 Rattus no

ALIGNMENTS

RESULT 1
LOCUS HUMTHRR 3472 bp mRNA linear PRI 03-AUG-1993
DEFINITION Human thrombin receptor mRNA, complete cds.
ACCESSION M62424
VERSION M62424.1 GI:339676
KEYWORDS thrombin receptor.
SOURCE Human DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3472)
Vu,T.K., Hung,D.T., Wheaton,V.I. and Coughlin,S.R.
Molecular cloning of a functional thrombin receptor reveals a novel
proteolytic mechanism of receptor activation

JOURNAL Cell 64 (6), 1057-1068 (1991)
MEDLINE 91168254
PUBMED 1672365
FEATURES source Location/Qualifiers
1..3472
/organism="Homo sapiens"
/db_xref="taxon:9606"
225..1502
CDS
/codon_start=1
/product="thrombin receptor"
/protein_id="AAA36743.1"
/db_xref="GI:339677"

BC002464.1 GI:12803296
MGC.
Homo sapiens.
Homo sapiens
1 (bases 1 to 3479)
Strausberg, R.
Direct Submission
Submitted (05-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@hgrl.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Masello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tiongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Image: 5 Row: p Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 6031164.
Location/Qualifiers
1..3479
/organism="Homo sapiens"
/db_xref="LocusID:2149"
/db_xref="taxon:9606"
/clone="MGC:1197 IMAGE:3343051"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_16"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
190..1467
/codon_start=1
/product="coagulation factor II (thrombin) receptor"
/protein_id="AAH02464.1"
/db_xref="GI:12803297"
/translation="MGPRLLIIAACFSLGCLLSARTRARPESKATNATLDPRSF
LRPNDRKPEFWEDEKESGLFEYRLVSLINKSPLOKQLPAPISDASGLTSSWLT
LFVPSVTVGVVSLPLNIAIVFILMKVKPAPVYIMHLATADVLVSVLPKIS
LYFGSDQWQFSELRCVTAAYFCNMYSITLLMTVIDRFLAVYPMOSLSWRTLGR
ASTFCLAIWALATAGVVPVLLKQITQVPLNITTCVDVNLLEGGYAYYFSAFSA
VFEPVPLIISTVCVSTIRCLSSAVANRKSRAFLSAAVFCIFIIICGPTNVLII
AHYFSLSHSTTTEAAYFAYLLCVSVSSICDPLIYVYASSECQRYVYSILCKRESS
DPSYNSGQGMASKMDTCSNLSNLSYKKLLI"

BASE COUNT 933 a 817 c 785 g 937 t
ORIGIN
Query Match 100.0%; Score 548; DB 9; Length 3472;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGCCGAGGTCGCTTGGACCTGTATCTACCGTGGGACCCCTGCGCTCTCCCTGCCGCG 60
DB 64 CGCCGAGGTCGCTTGGACCTGTATCTACCGTGGGACCCCTGCGCTCTCCCTGCCGCG 123
QY 61 AGACCGGTCCTCCGAGCAGCCGCGAGAGGTGAGGAGAGGTGAGGAGCAGCGAGCCGAGGC 120
DB 124 AGACCGGTCCTCCGAGCAGCCGCGAGAGGTGAGGAGAGGTGAGGAGCAGCGAGCCGAGGC 183
QY 121 GGGCAGCCTCCCGAGCAGCCGCGAGCAGCCGCGAGAGGTGAGGAGAGGTGAGGAGCAGCGAGGC 180
DB 184 GGGCAGCCTCCCGAGCAGCCGCGAGCAGCCGCGAGAGGTGAGGAGAGGTGAGGAGCAGCGAGGC 243
QY 181 TGCTGTGTCGCGCGCTTCACTGTGTCGCGCGCGCTGTGTCTGCCGCGACCCGCGGCC 240
DB 244 TGCTGTGTCGCGCGCTTCACTGTGTCGCGCGCGCTGTGTCTGCCGCGACCCGCGGCC 303
QY 241 GCAGCCAGCAATCAAAAGCAACAAATGCCACCTTAGATCCCGGTCATTTCTTCTCAGGA 300
DB 304 GCAGCCAGCAATCAAAAGCAACAAATGCCACCTTAGATCCCGGTCATTTCTTCTCAGGA 363
QY 301 ACCCCCAATGATAATATGACCAATTTTGGGAGGATGAGGAGAAATGAAAGTGGGTAA 360
DB 364 ACCCCCAATGATAATATGACCAATTTTGGGAGGATGAGGAGAAATGAAAGTGGGTAA 423
QY 361 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAACAACTTCTGTCAT 420
DB 424 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAACAACTTCTGTCAT 483
QY 421 TCATCTCAGAAGATGCTCCGGATATTGACACAGTCTCTGCGTGACACTTTTGTGCCAT 480
DB 484 TCATCTCAGAAGATGCTCCGGATATTGACACAGTCTCTGCGTGACACTTTTGTGCCAT 543
QY 481 CTGTGTACACCGGAGTGTGTAGTCAGCTCCCACTAAACATCATGGCCATGTTGTGT 540
DB 544 CTGTGTACACCGGAGTGTGTGTAGTCAGCTCCCACTAAACATCATGGCCATGTTGTGT 603
QY 541 TCATCTCTG 548
DB 604 TCATCTCTG 611

RESULT 2
BC002464
LOCUS
DEFINITION
Homo sapiens, coagulation factor II (thrombin) receptor, clone
MGC:1197 IMAGE:3343051, mRNA, complete cds.
3479 bp mRNA linear PRI 12-JUL-2001

Query Match 100.0%; Score 548; DB 9; Length 3479;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGCCGAGGTCGCTTGGACCTGTATCTACCGTGGGACCCCTGCGCTCTCCCTGCCGCG 60
DB 29 CGCCGAGGTCGCTTGGACCTGTATCTACCGTGGGACCCCTGCGCTCTCCCTGCCGCG 88
QY 61 AGACCGGTCCTCCGAGCAGCCGCGAGAGGTGAGGAGAGGTGAGGAGCAGCGAGCCGAGGC 120

[illegible]

```
QY 301 ACCCCATGATAAATATGACCAATTTTGGAGGATGAGGAGAAAATGAAAGTGGGTTAA 360
Db 364 ACCCCATGATAAATATGAAACCAATTTTGGAGGATGAGGAGAAAATGAAAGTGGGTTAA 423
QY 361 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAAT 420
Db 424 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAAT 483
QY 421 TCATCTCAGAAGATGCTCCGGATATTTGACACGCTCCGCTGACACATCTTTGTCCCAT 480
Db 484 TCATCTCAGAAGATGCTCCGGATATTTGACACGCTCCGCTGACACATCTTTGTCCCAT 543
QY 481 CTGTGTACACGGAGTCTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
Db 544 CTGTGTACACGGAGTCTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603
QY 541 TCATCCTG 548
Db 604 TCATCCTG 611

RESULT 5
AR065265
LOCUS AR065265 3480 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 219 from patent US 5849507.
ACCESSION AR065265
VERSION AR065265.1 GI:5995481
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3480)
AUTHORS Coughlin,S.R.
TITLE Methods to diagnose thrombosis by measuring activation peptide
JOURNAL Patent: US 5849507-A 219 15-DEC-1998;
FEATURES
Location/Qualifiers
1..3480
/organism="unknown"
BASE COUNT 941 a 817 c 785 g 937 t
ORIGIN
Query Match 100.0%; Score 548; DB 6; Length 3480;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCGAGGGTGGACCTGATCTTACCCGTGGGACCCCTGGCTCTCGCTGCGCG 60
Db 64 CGCCGAGGGTGGCTTGGACCCCTGATCTTACCCGTGGGACCCCTGGCTCTCGCTGCGCG 123
QY 61 AAGACCGGCTCCCGACCCGCAAGTCAAGAGAGAGGGTGAAGCGGAGAGCCCGAGGC 120
Db 124 AAGACCGGCTCCCGACCCGCAAGTCAAGAGAGAGGGTGAAGCGGAGAGCCCGAGGC 183
QY 121 GGGGAGCGCTCCCGAGAGCGCCGCGAGAGCCGCGGACAAATGGGCGCGCGGCTGC 180
Db 184 GGGGAGCGCTCCCGAGAGCGCCGCGAGAGCCGCGGACAAATGGGCGCGCGGCTGC 243
QY 181 TGCTGTGGCGCGCTGCTTCACTGTGTGGCGCCGCTGTGTCTGCCCGCACCCGGGCCC 240
Db 244 TGCTGTGGCGCGCTGCTTCACTGTGTGGCGCCGCTGTGTCTGCCCGCACCCGGGCCC 303
QY 241 GCAGGCCAGAAATCAAAAGCAACAATGCCACCTTAGATCCCGGGTCAATTTCTTCAGGA 300
Db 304 GCAGGCCAGAAATCAAAAGCAACAATGCCACCTTAGATCCCGGGTCAATTTCTTCAGGA 363
QY 301 ACCCCATGATAAATGAAACCAATTTTGGAGGATGAGGAGAAAATGAAAGTGGGTTAA 360
Db 364 ACCCCATGATAAATGAAACCAATTTTGGAGGATGAGGAGAAAATGAAAGTGGGTTAA 423
QY 421 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAAT 420
Db 424 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAAT 483
QY 481 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAAT 540
Db 424 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAAT 603
```

```
QY 421 TCATCTCAGAAGATGCTCCCGATATTTGACACGCTCTCGCTGACACTCTTTGTCCCAT 480
Db 484 TCATCTCAGAAGATGCTCCCGATATTTGACACGCTCTCGCTGACACTCTTTGTCCCAT 543
QY 481 CTGTGTACACGGAGTCTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
Db 544 CTGTGTACACGGAGTCTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603
QY 541 TCATCCTG 548
Db 604 TCATCCTG 611

RESULT 6
AR137610
LOCUS AR137610 3480 bp DNA linear PAT 16-JUN-2001
DEFINITION Sequence 219 from patent US 6197541.
ACCESSION AR137610
VERSION AR137610.1 GI:14479119
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 3480)
AUTHORS Coughlin,S.R.
TITLE Recombinant thrombin receptors and assays using them
JOURNAL Patent: US 6197541-A 219 06-MAR-2001;
FEATURES
Location/Qualifiers
1..3480
/organism="unknown"
BASE COUNT 941 a 817 c 785 g 937 t
ORIGIN
Query Match 100.0%; Score 548; DB 6; Length 3480;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCCGAGGGTGGCTTGGACCCCTGATCTTACCCGTGGGACCCCTGGCTCTCGCTGCGCG 60
Db 64 CGCCGAGGGTGGCTTGGACCCCTGATCTTACCCGTGGGACCCCTGGCTCTCGCTGCGCG 123
QY 61 AAGACCGGCTCCCGACCCGCAAGTCAAGAGAGAGGGTGAAGCGGAGAGCCCGAGGC 120
Db 124 AAGACCGGCTCCCGACCCGCAAGTCAAGAGAGAGGGTGAAGCGGAGAGCCCGAGGC 183
QY 121 GGGGAGCGCTCCCGAGAGCGCCGCGAGAGCCGCGGACAAATGGGCGCGCGGCTGC 180
Db 184 GGGGAGCGCTCCCGAGAGCGCCGCGAGAGCCGCGGACAAATGGGCGCGCGGCTGC 243
QY 181 TGCTGTGGCGCGCTGCTTCACTGTGTGGCGCCGCTGTGTCTGCCCGCACCCGGGCCC 240
Db 244 TGCTGTGGCGCGCTGCTTCACTGTGTGGCGCCGCTGTGTCTGCCCGCACCCGGGCCC 303
QY 241 GCAGGCCAGAAATCAAAAGCAACAATGCCACCTTAGATCCCGGGTCAATTTCTTCAGGA 300
Db 304 GCAGGCCAGAAATCAAAAGCAACAATGCCACCTTAGATCCCGGGTCAATTTCTTCAGGA 363
QY 301 ACCCCATGATAAATGAAACCAATTTTGGAGGATGAGGAGAAAATGAAAGTGGGTTAA 360
Db 364 ACCCCATGATAAATGAAACCAATTTTGGAGGATGAGGAGAAAATGAAAGTGGGTTAA 423
QY 361 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAAT 420
Db 424 CTGAATACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAAT 483
QY 421 TCATCTCAGAAGATGCTCCCGATATTTGACACGCTCTCGCTGACACTCTTTGTCCCAT 480
Db 484 TCATCTCAGAAGATGCTCCCGATATTTGACACGCTCTCGCTGACACTCTTTGTCCCAT 543
QY 481 CTGTGTACACGGAGTCTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 540
Db 544 CTGTGTACACGGAGTCTTTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGT 603
```


Qy 541 TCATCCTG 548
|||||||
Db 604 TCATCCTG 611

RESULT 7
LOCUS I74660
DEFINITION Sequence 219 from patent US 5688768.
ACCESSION I74660
VERSION I74660.1 GI:3010801
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 3480)
AUTHORS Coughlin,S.R. and Scarborough,R.M.
TITLE Recombinant thrombin receptor and related pharmaceuticals
JOURNAL Patent: US 5688768-A 219 18-NOV-1997;
FEATURES
Location/Qualifiers
1..3480
source /organism="unknown"
BASE COUNT 941 a 817 c 785 g 937 t
ORIGIN

Query Match 100.0%; Score 548; DB 6; Length 3480;
Best Local Similarity 100.0%; Pred. No. 3.3e-112;
Matches 548; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCCGAGGTCGCTTGGACCTGATCTACCGTGGCACCCTGCGCTCTGCTGCGCG 60
Db 64 CGCCGAGGTCGCTTGGACCTGATCTACCGTGGCACCCTGCGCTCTGCTGCGCG 123

Qy 61 AAGACCGGCTCCCGACCCGACAGATCAGAGAGAGGGTGAACCGAGCAGCCGAGGC 120
Db 124 AAGACCGGCTCCCGACCCGACAGATCAGAGAGAGGGTGAACCGAGCAGCCGAGGC 183

Qy 121 GGGGAGCCTCCCGGAGCAGCGCGGACAGCCCGGACAAATGGGCGCGCGGCTGC 180
Db 184 GGGGAGCCTCCCGGAGCAGCGCGGACAGCCCGGACAAATGGGCGCGCGGCTGC 243

Qy 181 TGTGCTGGCGGCTGCTCAGTCTGCGGCGCGCTGTGCTGCGCGACCCGCGGCC 240
Db 244 TGTGCTGGCGGCTGCTCAGTCTGCGGCGCGCTGTGCTGCGCGACCCGCGGCC 303

Qy 241 CGAGCCGAGATCAAAAGCAACAAATAGACCTTAGATCCCGGTCATTTCTTCTCAGGA 300
Db 304 CGAGCCGAGATCAAAAGCAACAAATAGACCTTAGATCCCGGTCATTTCTTCTCAGGA 363

Qy 301 ACCCCCAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 364 ACCCCCAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 423

- Qy 361 CTGAATACAGATAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAT 420
Db 424 CTGAATACAGATAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACATTCCTGCAT 483

Qy 421 TCATCTCAGAAGATGCTCCGATATTTGACCAAGTCTCTGCTGACACTTTTGTCCCAT 480
Db 484 TCATCTCAGAAGATGCTCCGATATTTGACCAAGTCTCTGCTGACACTTTTGTCCCAT 543

Qy 481 CTGCTACACCGAGTGTGTTAGTACAGCTCCCACTTAAACATCATGGCCATCGTTGTGT 540
Db 544 CTGCTACACCGAGTGTGTTAGTACAGCTCCCACTTAAACATCATGGCCATCGTTGTGT 603

Qy 541 TCATCCTG 548
|||||||
Db 604 TCATCCTG 611

RESULT 8
AF028727
LOCUS

Papio hamadryas thrombin receptor (bthr12) mRNA, complete cds.
AF028727
VERSION AF028727.1 GI:2605876
KEYWORDS
SOURCE Papio hamadryas.
ORGANISM Papio hamadryas.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae; Papio.
1 (bases 1 to 3764)
Hayzer,D.J., Shoji,M. and Hanson,S.R.
cDNAs encoding the baboon thrombin receptor indicate a primate transcription start site upstream of putative sites reported for the human gene
Thromb. Res. 98 (2), 195-201 (2000)
20180190
PUBMED 10713321
2 (bases 1 to 3764)
Shoji,M., Hayzer,D.J. and Hanson,S.R.
Direct Submission
Submitted (07-OCT-1997) Hematologic Diseases Division, Building 1, CDC/NCID/DHA, Rm. 1310, Mail Stop D-02, 1600 Clifton Road, Atlanta, GA 30333, USA
Location/Qualifiers
1..3764
/organism="Papio hamadryas"
/db_xref="taxon:9557"
1..3764
/gene="bthr12"
532..1809
/gene="bthr12"
/codon_start=1
/product="thrombin receptor"
/protein_id="AAB84191.1"
/db_xref="GI:2605877"
/translation="MGPRRLILVAACILCGLPLLSARTRARRPASKATNATLDPRSFL
LRPNDKYEPFNEDEKNEKSLTEYRLVSNKSPLOKPLPAISDASGYLTSWLT
LFVPSVYTGTVFVSLPVMIAIVFLLKMKRKPAPVYVLMHATADVLFVSLPFLKIS
YILSGSDWQFSGELCRFVTAFAFCNNYASILLMTVISIDRFLLAVVPMOSLSWRTILGR
ASFTCLAIWALAIAGVPLLEKQITQVFLNITTDHVLNLTLLGYTYAYFSASA
VFFVFLIISTVCYVSIIRCLSSSTVANRKSRAFLSAAVFCIFICFGPTNILLI
AIYSELSTSTTAAYFAYLLCVCVSSISCCIDPLIYYVASSECQRYVYSILCKESS
DPSSNSNGQLMASKMDTCCSNLNSIYKLLT"
BASE COUNT 992 a 868 c 863 g 1041 t
ORIGIN

Query Match 80.6%; Score 441.6; DB 9; Length 3764;
Best Local Similarity 91.8%; Pred. No. 2e-88;
Matches 503; Conservative 0; Mismatches 34; Indels 11; Gaps 3;

Qy 1 CGCCGAGGTCGCTTGGACCTGATCTTACCCTGGGACCCCTGCGCTGCTGCGCG 60
Db 382 CGCCGGGGTGGCTCGGACCCCTGCTTGGCCCGG-----CGCGTGGCTGCGCG 432

Qy 61 AAGACCGGCTCCCGGACCCGACAGAGTCAAGAGAGAGGGTGAAGCGAGCAGCCCGAGGC 120
Db 433 AAGACCGGCTCCCGGACCCGACAGAGTCAAGAGAGAGGGTGAAGCGAGCAGCCCGAGGC 491

Qy 121 GGGGAGCCTCCCGGAGCAGCGCGCGAGAGCCCGGACAAATGGGCGCGCGCGCTGC 180
Db 492 GGGG- AACCTCTCTGGAGCAGCGCAGTCCCGGACAAATGGGCGCGCGCGCTGC 550

Qy 181 TGTGCTGGCGGCTGCTTCAAGTCTGCGGCGCGCTGTGCTGCGCGCGCGCGCGCG 240
Db 551 TGTGCTGGCGGCTGCTTCAAGTCTGCGGCGCGCTGTGCTGCGCGCGCGCGCGCG 610

Qy 241 GCAGCCAGATCAAAAGCAACAAATGCCACCTTAGATCCCGGTCATTTCTTCTCAGGA 300
Db 611 GCAGCCAGATCAAAAGCAACAAATGCCACCTTAGATCCCGGTCATTTCTTCTCAGGA 670

Qy 301 ACCCCCAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 671 ATCCCAATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 730

* and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 849: contig of 849 bp in length
 850 949: gap of 100 bp
 950 1768: contig of 819 bp in length
 1769 1868: gap of 100 bp
 1869 2679: contig of 811 bp in length
 2680 2779: gap of 100 bp
 2780 3583: contig of 804 bp in length
 3584 3683: gap of 100 bp
 3684 4465: contig of 782 bp in length
 4466 4565: gap of 100 bp
 4566 5401: contig of 836 bp in length
 5402 5501: gap of 100 bp
 5502 6340: contig of 839 bp in length
 6341 6440: gap of 100 bp
 6441 7270: contig of 830 bp in length
 7271 7370: gap of 100 bp
 7371 8217: contig of 847 bp in length
 8218 8317: gap of 100 bp
 8318 9141: contig of 824 bp in length
 9142 9241: gap of 100 bp
 9242 10057: contig of 816 bp in length
 10058 10157: gap of 100 bp
 10158 10986: contig of 829 bp in length
 10987 11086: gap of 100 bp
 11087 11911: contig of 825 bp in length
 11912 12011: gap of 100 bp
 12012 12788: contig of 777 bp in length
 12789 12888: gap of 100 bp
 12889 13702: contig of 814 bp in length
 13703 13802: gap of 100 bp
 13803 14634: contig of 832 bp in length
 14635 14734: gap of 100 bp
 14735 15571: contig of 837 bp in length
 15572 15671: gap of 100 bp
 15672 16504: contig of 833 bp in length
 16505 17393: contig of 789 bp in length
 17394 17493: gap of 100 bp
 17494 18327: contig of 834 bp in length
 18328 18427: gap of 100 bp
 18428 19259: contig of 832 bp in length
 19260 19359: gap of 100 bp
 19360 20208: contig of 849 bp in length
 20209 20308: gap of 100 bp
 20309 21141: contig of 833 bp in length
 21142 21241: gap of 100 bp
 21242 22060: contig of 819 bp in length
 22061 22160: gap of 100 bp
 22161 22966: contig of 806 bp in length
 22967 23066: gap of 100 bp
 23067 23869: contig of 803 bp in length
 23870 23969: gap of 100 bp
 23970 24798: contig of 829 bp in length
 24799 24898: gap of 100 bp
 24899 25702: contig of 804 bp in length
 25703 25802: gap of 100 bp
 25803 26567: contig of 765 bp in length
 26568 26667: gap of 100 bp
 26668 27483: contig of 816 bp in length
 27484 27583: gap of 100 bp
 27584 28412: contig of 829 bp in length
 28413 28512: gap of 100 bp
 28513 29344: contig of 832 bp in length
 29345 29444: gap of 100 bp
 29445 30271: contig of 827 bp in length

Qy 361 CTGAATACAGATTAGTCTCCATCAATAAAAAAGCAGTCTCTTTCAAAAACAACTTCCTGCAT 420
 Db 731 CTGAATACAGATTAGTCTCCATCAATAAAAAAGCAGTCTCTTTCAAAAACAACTTCCTGCAT 790
 Qy 421 TCATCTCAGAAGATGCTCCGGATATTTGACACGCTCCGCTGACACACTTTGTGCCAT 480
 Db 791 TCATCTCAGAAGATGCTCCGGATATTTGACACGCTCCGCTGACACACTTTGTGCCAT 850
 Qy 481 CTGTGTACACGGAGTGTGTGTAGTACGCTCCCACTAAACATCATGGCCATGCTGTGTGT 540
 Db 851 CTGTGTACACGGAGTGTGTGTAGTACGCTCCCGGTAACATCATGGCCATGCTGTGTGT 910
 Qy 541 TCATCCTG 548
 Db 911 TCATCCTG 918

RESULT 9
 AC026151 70679 bp DNA linear HTG 19-MAR-2000
 . LOCUS Homo sapiens clone RP11-329A20, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION
 AC026151
 VERSION
 AC026151.1 GI:7264221
 KEYWORDS
 HTG; HTGS, PHASE0.
 SOURCE
 Homo sapiens.

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens, clone RP11-329A20
 Unpublished

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 1 (bases 1 to 70679)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Glnde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., LaRoque, K., Lanazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
 Vassiliev, H., Vleel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission

TITLE
 JOURNAL
 COMMENT
 Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: l8625
 Center clone name: 329_A_20

* NOTE: This record contains 77 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads

* 30272 30371: gap of 100 bp
* 30372 31143: contig of 772 bp in length
* 31144 31243: gap of 100 bp
* 31244 32053: contig of 810 bp in length
* 32054 32153: gap of 100 bp
* 32154 32962: contig of 809 bp in length
* 32963 33062: gap of 100 bp
* 33063 33883: contig of 821 bp in length
* 33884 33983: gap of 100 bp
* 33984 34807: contig of 824 bp in length
* 34808 34907: gap of 100 bp
* 34908 35712: contig of 805 bp in length
* 35713 35812: gap of 100 bp
* 35813 36643: contig of 831 bp in length
* 36644 36743: gap of 100 bp
* 36744 37572: contig of 829 bp in length
* 37573 37672: gap of 100 bp
* 37673 38499: contig of 827 bp in length
* 38500 38599: gap of 100 bp
* 38600 39425: contig of 826 bp in length
* 39426 39525: gap of 100 bp
* 39526 40305: contig of 780 bp in length
* 40306 40405: gap of 100 bp
* 40406 41233: contig of 828 bp in length
* 41234 41333: gap of 100 bp
* 41334 42152: contig of 819 bp in length
* 42153 42252: gap of 100 bp
* 42253 43091: contig of 839 bp in length
* 43092 43191: gap of 100 bp
* 43192 44035: contig of 844 bp in length
* 44036 44135: gap of 100 bp
* 44136 44973: contig of 838 bp in length
* 44974 45073: gap of 100 bp
* 45074 45888: contig of 815 bp in length
* 45889 45988: gap of 100 bp
* 45989 46837: contig of 849 bp in length
* 46838 46937: gap of 100 bp
* 46938 47765: contig of 828 bp in length
* 47766 47865: gap of 100 bp
* 47866 48665: contig of 800 bp in length
* 48666 48765: gap of 100 bp
* 48766 49576: contig of 811 bp in length
* 49577 49676: gap of 100 bp
* 49677 50500: contig of 824 bp in length
* 50501 50600: gap of 100 bp
* 50601 51437: contig of 837 bp in length
* 51438 51537: gap of 100 bp
* 51538 52369: contig of 832 bp in length
* 52370 52469: gap of 100 bp
* 52470 53296: contig of 827 bp in length
* 53297 53396: gap of 100 bp
* 53397 54202: contig of 806 bp in length
* 54203 54302: gap of 100 bp
* 54303 55128: contig of 826 bp in length
* 55129 55228: gap of 100 bp
* 55229 56078: contig of 850 bp in length
* 56079 56178: gap of 100 bp
* 56179 57000: contig of 822 bp in length
* 57001 57100: gap of 100 bp
* 57101 57897: contig of 797 bp in length
* 57898 57997: gap of 100 bp
* 57998 58620: contig of 823 bp in length
* 58621 58920: gap of 100 bp
* 58921 59718: contig of 798 bp in length
* 59719 59818: gap of 100 bp
* 59819 60609: contig of 791 bp in length
* 60610 60709: gap of 100 bp
* 60710 61523: contig of 820 bp in length
* 61530 61629: gap of 100 bp
* 61630 62460: contig of 831 bp in length
* 62461 62560: gap of 100 bp
* 62561 63354: contig of 794 bp in length
* 63355 63454: gap of 100 bp

* 63455 64251: contig of 797 bp in length
Query Match 55.9%; Score 306.2; DB 2; Length 70679;
Best Local Similarity 69.6%; Pred. No. 3.9e-58;
Matches 332; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 72 CCCGACCCGAGAGTCAGGAGAGAGGGTGAAGCGGAGCAGCCGCGGAGCGGCCTC 131
Db 51413 CCAGACAGATAAGAGATGAGGAGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN 51472
QY 132 CCGGAGCAGCGCCGCGAGAGCCCGGACAAATGGGGCGCGGCTGCTGCTGGTGCC 191
Db 51473 NNN 51532
QY 192 GCCTGCTTCAGTCTGTGGCGCCCGCTGTGTCTCCGCGCACCCGCGGCGCGAGCA 251
Db 51533 NNNNNGCACACTTGTGCTGACCTCGAGGTTTTTCTCTAGNAGNATCCCTCCCTC 51592
QY 252 TCAAAAGCAACAATGCCACCTTAGATCCCGGTCATTTCTTCAGGAACCCCAATGAT 311
Db 51593 TCAAAAGCAACAATGCCACCTTAGATCCCGGTCATTTCTTCAGGAACCCCAATGAT 51652
QY 312 AAATATGAACCATTTTGGGAGGATGAGGAGAAAATGAAAGTGGTTAACTGAATACAGA 371
Db 51653 AAATATGAACCATTTTGGGAGGATGAGGAGAAAATGAAAGTGGTTAACTGAATACAGA 51712
QY 372 TTAGTCTCCATCAATAAAGCAGTCCTTCAAAAACAACCTTCCTGCATCATCTCAGAA 431
Db 51713 TTAGTCTCCATCAATAAAGCAGTCCTTCAAAAACAACCTTCCTGCATCATCTCAGAA 51772
QY 432 GATGCTCCGGATATTGACCACTCTGCTGCTGACACTCTTTGCCCATCTGTGTACACC 491
Db 51773 GATGCTCCGGATATTGACCACTCTGCTGCTGACACTCTTTGCCCATCTGTGTACACC 51832
QY 492 GGAGTGTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTGTGGTTCACTCCTG 548
Db 51833 GGAGTGTGTAGTCAGCCTCCCACTAAACATCATGGCCATCGTGTGGTTCACTCCTG 51889
RESULT 10
AF391809
LOCUS
DEFINITION Homo sapiens coagulation factor II (thrombin) receptor (F2R) gene,
complete cds.
ACCESSION AF391809
VERSION AF391809.2 GI:14971463
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 24771)
AUTHORS Rieder,M.J., Carrington,D.P., Chung,M.-W., Lee,K.L., Poel,C.L.,
Yi,Q. and Nickerson,D.A.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2001) Molecular Biotechnology, University of
Washington, 1705 NE Pacific, Seattle, WA 98195, USA
REFERENCE 2 (bases 1 to 24771)
AUTHORS Rieder,M.J. and Nickerson,D.A.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2001) Molecular Biotechnology, University of
Washington, 1705 NE Pacific, Seattle, WA 98195, USA
REMARK Sequence update by submitter
COMMENT On Jul 20, 2001 this sequence version replaced gi:14583134.
To cite this work please use: SeattleSNPs. NHLBI Program for
Genomic Applications, UW-FHCRC, Seattle, WA (URL:
http://pga.mbt.washington.edu).
FEATURES
Location/Qualifiers
Source 1. 24771
/organism="Homo sapiens"
/db_xref="taxon:9606"
repeat_region 55. .211
/rpt_family="MER104"


```
/frequency="0.01"
/replace="C"
5276
/genes="F2R"
/frequency="0.03"
/replace="A"
5428
/genes="F2R"
/frequency="0.01"
/replace="A"
5472
/genes="F2R"
/frequency="0.40"
/replace="A"
5480
/genes="F2R"
/frequency="0.01"
/replace="A"
5515
/genes="F2R"
/frequency="0.01"
/replace="A"

Query Match      55.1%; Score 302; DB 9; Length 24771;
Best Local Similarity 100.0%; Pred. No. 3.3e-57;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 CAGAATCAAAAGCAACAATGCCACCTAGATCCCGGTCAATTCCTCTCAGGAACCCCA 306
Db 18153 CAGAAATCAAAAGCAACAATGCCACCTAGATCCCGGTCAATTCCTCTCAGGAACCCCA 18212

QY 307 ATGATAAATATGAACCAATTTGGGAGGATGAGGAGAAAATGAAAGTGGGTTAACTGAAT 366
Db 18213 ATGATAAATATGAACCAATTTGGGAGGATGAGGAGAAAATGAAAGTGGGTTAACTGAAT 18272

QY 367 ACAGATTAGTCTCCATCAATATAAAGCAGTCCTCTTCAAAAACAACCTTCCTGCATTCATCT 426
Db 18273 ACAGATTAGTCTCCATCAATATAAAGCAGTCCTCTTCAAAAACAACCTTCCTGCATTCATCT 18332

QY 427 CAGAAGATGCTCCGGATATTGACCAAGTCCTGCTGCTGACACTCTTGTGCCATCTGTGT 486
Db 18333 CAGAAGATGCTCCGGATATTGACCAAGTCCTGCTGCTGACACTCTTGTGCCATCTGTGT 18392

QY 487 ACACGGAGTGTGTGAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTTCATCC 546
Db 18393 ACACGGAGTGTGTGAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTTCATCC 18452

QY 547 TG 548
Db 18453 TG 18454

RESULT 11
-AC020927
LOCUS      110965 bp DNA linear HTG 19-APR-2001
DEFINITION Homo sapiens chromosome 5 clone CTD-2153E7, WORKING DRAFT SEQUENCE,
4 ordered pieces.
ACCESSION AC020927
VERSION AC020927.4 GI:7711631
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 110965)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 110965)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
```

COMMENT

```
On May 6, 2000 this sequence version replaced gi:7417650.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 686404
Center clone name: CITB-H1_2153E7
-----
Summary Statistics
Consensus quality: 106980 bases at least Q40
Consensus quality: 109994 bases at least Q30
Consensus quality: 110455 bases at least Q20
Estimated insert size: 110000; pulse field gel estimation
Quality coverage: 7.03 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.98 in Q20 bases; pulse field gel estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 8728: contig of 8728 bp in length
* 8729 8828: gap of unknown length
* 11329 11329: contig of 2501 bp in length
* 11330 11429: gap of unknown length
* 11430 40589: contig of 29160 bp in length
* 40590 40689: gap of unknown length
* 40690 110965: contig of 70276 bp in length.

FEATURES
         Location/Qualifiers
         1..110965
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="CTD-2153E7"
            /clone_lib="CalTech human BAC library D"
BASE COUNT 32294 a 23208 c 24149 g 31014 t 300 others
ORIGIN
```

```
Query Match      55.1%; Score 302; DB 2; Length 110965;
Best Local Similarity 100.0%; Pred. No. 3.4e-57;
Matches 302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 247 CAGAATCAAAAGCAACAATGCCACCTAGATCCCGGTCAATTCCTCTCAGGAACCCCA 306
Db 44895 CAGAATCAAAAGCAACAATGCCACCTAGATCCCGGTCAATTCCTCTCAGGAACCCCA 44954

QY 307 ATGATAAATATGAACCAATTTGGGAGGATGAGGAGAAAATGAAAGTGGGTTAACTGAAT 366
Db 44955 ATGATAAATATGAACCAATTTGGGAGGATGAGGAGAAAATGAAAGTGGGTTAACTGAAT 45014

QY 367 ACAGATTAGTCTCCATCAATATAAAGCAGTCCTCTTCAAAAACAACCTTCCTGCATTCATCT 426
Db 45015 ACAGATTAGTCTCCATCAATATAAAGCAGTCCTCTTCAAAAACAACCTTCCTGCATTCATCT 45074

QY 427 CAGAAGATGCTCCGGATATTGACCAAGTCCTGCTGCTGACACTCTTGTGCCATCTGTGT 486
Db 45075 CAGAAGATGCTCCGGATATTGACCAAGTCCTGCTGCTGACACTCTTGTGCCATCTGTGT 45134

QY 487 ACACGGAGTGTGTGAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTTCATCC 546
Db 45135 ACACGGAGTGTGTGAGTCAGCCTCCCACTAAACATCATGGCCATCGTTGTGTTTCATCC 45194

QY 547 TG 548
Db 45195 TG 45196
```

RESULT 13
AC025188 143335 bp DNA linear PRI 12-DEC-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2384B11, complete sequence.
DEFINITION AC025188
ACCESSION AC025188.8 GI:17530707
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143335)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 143335)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-REFERENCE 3 (bases 1 to 143335)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 143335)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 12, 2001 this sequence version replaced gi:16328263.
Draft Sequence produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
FEATURES
source
1. 143335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2384B11"
BASE COUNT 42376 a 29536 c 30902 g 40521 t
ORIGIN
Query Match 55.18; Score 302; DB 9; Length 143335;
Best Local Similarity 100.0%; Pred. No. 3.4e-57; Indels 0; Gaps 0;
Matches 302; Conservative 0; Mismatches 0;
QY 247 CAGAATAAAGCAACAAATGCCACCTTAGATCCCGGTCAATTTCTTCTCAGGAACCCCA 306
Db 73894 CAGAATCAAAAGCACAAATGCCACCTTAGATCCCGGTCAATTTCTTCTCAGGAACCCCA 73953
QY 307 ATGATAAATGAACCAATTTGGGAGGATGAGGAGAAAATGAAAGTGGGTAACTGAAT 366
Db 73954 ATGATAAATGAACCAATTTGGGAGGATGAGGAGAAAATGAAAGTGGGTAACTGAAT 74013
QY 367 ACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACTTCTGCAATCACT 426
Db 74014 ACAGATTAGTCTCCATCAATAAAGCAGTCTCTTCAAAAACAACTTCTGCAATCACT 74073
QY 427 CAGAAGATCCCTCCGGATATTTGACAGCTCCTGGCTGACACTCTTTGTCCTCATCTGTGT 486
Db 74074 CAGAAGATCCCTCCGGATATTTGACAGCTCCTGGCTGACACTCTTTGTCCTCATCTGTGT 74133
QY 487 ACACCGGAGTGTGTAGTACACCTCCCACTAAACATCATGCGCATCGTTGTGTTCACTCC 546
Db 74134 ACACCGGAGTGTGTAGTACACCTCCCACTAAACATCATGCGCATCGTTGTGTTCACTCC 74193
QY 547 TG 548
Db 74194 TG 74195

RESULT 14
AC008933 88881 bp DNA linear HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2297A3, WORKING DRAFT SEQUENCE,
DEFINITION AC008933
ACCESSION AC008933.6 GI:13676982
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 88881)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 88881)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 19, 2001 this sequence version replaced gi:9964732.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
